PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12Q 1/68, C07H 19/00, 21/00

A1

(11) International Publication Number: WO 99/07890

(43) International Publication Date: 18 February 1999 (18.02.99)

(21) International Application Number: PCT/US98/16211

(22) International Filing Date: 5 August 1998 (05.08.98)

(30) Priority Data:
08/906,494 5 August 1997 (05.08.97) US

(71) Applicant: UNIVERSITY OF MASSACHUSETTS [US/US]; 26th floor, 1 Beacon Street, Boston, MA 02108 (US).

(72) Inventor: STERN, Seth; 18B Riverview Road, Sterling, MA 01564 (US).

(74) Agent: FASSE, J., Peter; Fish & Richardson P.C., 225 Franklin Street, Boston, MA 02110-2804 (US).

(81) Designated States: AU, CA, JP, KR, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(54) Title: NUCLEIC ACID ISOLATION, QUANTITATION, AND STRUCTURE PROBING

(57) Abstract

Hybridization-based methods and ligation-dependent methods for nucleic acid isolation and quantitation. Also featured are methods for nucleic acid structure probing and for studying nucleic acid-ligand interaction.

NUCLEIC ACID ISOLATION, QUANTITATION, AND STRUCTURE PROBING

Field of the Invention

The invention relates to isolation, quantitation, and structure probing of nucleic acids.

Background of the Invention

Chemical and enzymatic probes used in nucleic acid structure probing react with, and typically covalently modify, specific atoms in the bases or phosphodiester backbone of nucleic acids. Any given atom in a nucleic acid may or may not react with a probe, depending on the specificity of the probe as well as the accessibility and chemical environment of the atom in the nucleotide. For example, dimethyl sulfate (DMS) specifically methylates N1 of adenine (A), N3 of cytosine (C), and N7 of guanine (G), but only if those atoms are in a reactive state or reside within an appropriate structure. For example, N1 of A and N3 of C cannot be methylated if the two bases are engaged in Watson-Crick base-pairing, because these nitrogen atoms directly participate in base-pairing hydrogen bonds, which preclude DMS reaction. Thus, DMS reactivity at A and C reflects the base pairing state of these residues, and can provide information about the secondary structure of a nucleic acid molecule.

Similarly, reactive atoms in nucleic acids may lose all or part of their reactivity toward a modifying agent when the nucleic acids bind to their ligands. For instance, when a small molecule ligand is hydrogen-bonded with N1 of A, this nitrogen atom will become unreactive towards DMS. Thus, structure probing not only generates detailed information about the conformation of nucleic

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

acids, but can also provide insight to the ligand-binding state of specific atoms within nucleic acid molecules.

In some cases, an additional, post-modification treatment may be used to aid detection of a covalent modification by inducing backbone strand scissions at modified positions. For instance, methylation of N7 of G by DMS and carbethoxylation of N7 of A and G by diethyl pyrocarbonate ("DEPC") can be detected by creating a strand scission at the modified base by incubation with sodium borohydride and aniline (for RNA) or pyridine (for DNA).

Fig. 1 illustrates a conventional method of RNA structure probing using DMS as a probing agent. DMS modification disrupts base pairing between the modified template ribonucleotide and incoming deoxyribonucleotide triphosphates during reverse transcription, thereby creating complementary DNA ("cDNA") transcripts truncated at the modified nucleotides. The figure shows the transcripts (SEQ ID NOS:2-5, 7 and 8) generated by reverse transcription of a DMS-modified RNA (SEQ ID NO:1). Chemical modification conditions are typically adjusted so that, on average, no more than one modification (or truncation) occurs per RNA molecule. In Fig. 1, all DMS modifications are shown on a single RNA molecule for convenience. To permit reverse transcription of the entire RNA molecule, a heterologous single-stranded sequence serving as a priming site is added to the 3′ end of the RNA. Three potentially interesting truncated fragments, i.e., those corresponding to A1408 (SEQ ID NO:8), A1492 (SEQ ID NO:3), and A1493 (SEQ ID NO:4) (that is, fragments starting at the adenine at nucleotide locations 1408, 1492, and 1493), are generated. However, due to the low rate of modification, the reverse transcription also generates a substantial amount of full-length transcript

(SEQ ID NO:7). Thus, signals generated by any particular truncated fragment will be obscured by those generated by the full-length and other transcripts. This signal contamination problem is commonly solved by separation of the transcripts by denaturing polyacrylamide gel electrophoresis ("PAGE").

Summary of the Invention

The invention features methods for isolating or isolating and quantitating specific target nucleic acid fragments (e.g., DNA, RNA, or DNA/RNA hybrids) from mixtures of nucleic acid fragments.

In one of these methods (i.e., a hybridization-based selection method), a non-target fragment is removed by hybridizing it to an immobilized (an oligo can be initially immobilized or immobilized after hybridization) "subtraction oligonucleotide" (oligonucleotides are also referred to herein as "oligos") that is complementary to a known sequence present in the non-target fragment, but absent in the target fragment. This removal step is repeated to remove additional non-target fragments until the known sequence in the target fragment becomes unique among the remaining fragments. Then the target fragment is specifically selected by hybridizing it to an immobilized selection oligo that is complementary to the unique sequence.

A "unique sequence" is a sequence that enables the specific hybridization, and thereby selection, of a nucleic acid fragment in which it is contained.

The nucleic acid fragments, including the target, can be products of a transcription reaction and can also include a signal-producing agent (e.g., a radioactive isotope or a fluorophore). In the new method, regions of hybridization or complementarity between the oligos and their target sequences can contain

gaps, provided that a sufficient base-pairing interaction is maintained to permit isolation.

In another new method for nucleic acid isolation and quantitation (i.e., a ligation-dependent selection method), the target nucleic acid fragment has a unique, single-stranded terminal sequence. In this method, a double-stranded selection oligonucleotide is used. This oligo has (i) a first (i.e., top or upper) strand including a protruding portion that complements the unique terminal sequence of the target, and a non-protruding portion, and (ii) a second (i.e., bottom or lower) strand that is complementary to the non-protruding portion. If desired, the two strands can be covalently linked. The double-stranded region of the oligo can also be G/C rich to provide higher stability.

To isolate the target from a mixture of nucleic acid fragments in this method, the mixture is incubated with the oligo to allow hybridization between the protruding portion and the unique terminal sequence. The unique terminal sequence of the target is then ligated with the second strand of the oligo. Nucleic acid fragments that are not ligated to the second strand are removed, thereby separating the target fragment from non-target fragments. The amount of the recovered target fragment can be measured to quantitate the target fragment.

In the ligation-dependent selection method, the target fragment can contain a signal-producing agent, while the oligo is immobilized onto a solid surface via one or both of its strands. The signal retained on the solid surface after non-ligated fragments have been removed then corresponds to the amount of the target present in the mixture. If the oligo is immobilized via its first strand, non-ligated nucleic acid fragments can be removed by a semi-denaturing wash. This wash

denatures base-pairing between the protruding portion of the first strand and any other nucleotide sequence, but maintains base-pairing between the non-protruding portion of the first strand and the second strand in a region (e.g., 2 or more basepairs) of the oligo.

The oligo can be part of an oligo array in which each of the oligos is immobilized at a distinct and pre-determined location on a solid surface.

Alternatively, the nucleic acid fragments, including the target, can be immobilized, while the double-stranded oligo is labeled with a signal producing agent. If the first strand is labeled, non-ligated double-stranded oligos can be removed by a semi-denaturing wash as described above.

The new isolation/quantitation methods described above can be applied to nucleic acid structure probing, e.g., to determine whether a given nucleotide in a nucleic acid can be modified by a chemical or enzymatic probing (i.e., modifying) agent. To achieve this, the nucleic acid is incubated with the modifying agent under conditions that allow modification of the nucleic acid. Subsequently, nucleic acid fragments are generated from the incubated nucleic acid, where modification of the given nucleotide will result in a target nucleic acid fragment whose terminus corresponds to the given nucleotide. The target fragment is then isolated using the new isolation/quantitation methods, wherein the presence of the target fragment indicates that the given nucleotide is modified by the modifying agent. When the nucleic acid fragments comprise a signal-producing agent, the presence and amount of the target can be indicated by the signal detected after the target has been isolated.

To generate the nucleic acid fragments from the probed (or modified) nucleic acid, the probed nucleic acid can be contacted with a cleaving agent that cleaves

nucleic acids only at an appropriately modified nucleotide; alternatively, the probed nucleic acid can be transcribed to generate the nucleic acid fragments, where the transcription terminates at any appropriately modified nucleotide. If the probed nucleic acid is an RNA, it can be degraded after transcription with an RNase to facilitate subsequent isolation steps. If the probed nucleic acid is a DNA, the transcription step can be repeated by denaturing a nucleic acid duplex formed by the nucleic acid template and its transcript, and annealing the template with a primer for a new round of transcription; this will increase the yield of single-stranded nucleic acid fragments for subsequent isolation/quantitation.

The new structure probing methods can be used to determine whether a compound can alter the reactivity of a given nucleotide in a test nucleic acid toward a modifying agent. To do this, the test nucleic acid is incubated with the compound, and then the modifiability of the given nucleotide by the modifying agent is determined. A change in the modifiability of the given nucleotide following treatment with the compound indicates that the compound alters the reactivity of the given nucleotide toward the modifying agent. For instance, the compound will be found to increase the reactivity of the nucleotide if the given nucleotide is not modified without treatment of the compound and is modified following treatment of the compound. Conversely, the compound will be regarded as being capable of decreasing the nucleotide's reactivity if the nucleotide is modified without the compound treatment but is modified with the treatment.

The new methods are fully automatable. In particular, when used in nucleic acid probing, the new methods eliminate the need for PAGE, and therefore the

need for laborious human intervention (e.g., pouring, mounting, loading, running, and disassembling polyacrylamide gel). Thus, the new structure probing methods can be readily adapted for high-throughput nucleic acid structure probing and compound screenings.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, exemplary methods and materials are described below. All publications, including patent applications and patents, and other references mentioned herein are incorporated by reference in their entirety. In case of conflict, the present specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

Other features or advantages of the present invention will be apparent from the following drawing and detailed description, and also from the appending claims.

Brief Description of the Drawings

Fig. 1 is a schematic diagram illustrating a conventional method for RNA structure probing.

Figs. 2a-2d are schematic diagrams illustrating a substraction-selection method for isolating a cDNA transcript of interest (SEQ ID NO:8) from a mixture of reverse transcription products. Poly(N) represents the primer (SEQ ID NO:5) for reverse transcription, which is end-labeled with ^{32}P; "<>" represents biotin-streptavidin coupling; "Bead" and "Well" each represent solid surfaces; and "supe" refers to supernatant. Four exemplary cDNA transcripts (SEQ ID NOS:3, 4, 7, and 8)

are shown, with the one truncated at A1408 designated as the transcript of interest (SEQ ID NO:8).

Fig. 3 is a diagram illustrating a ligation-mediated selection method. Symbols are the same as those described above for Fig. 2.

Fig. 4 is a diagram illustrating use of a ligation-mediated selection method (i.e., a differential labeling method) for simultaneous quantitation of 4 target nucleic acids. "Bio" stands for biotin; "St" stands for streptavidin; "Fluor$_n$" (n = 1-4) stands for a fluorophore emitting at a distinct wavelength.

Figs. 5a and 5b are lists of sequences (SEQ ID NOS:15-43) used in an experiment that characterized sequence specificity of ligation.

Fig. 6 is a schematic diagram illustrating a new ligation-dependent selection method in which the selection oligo is immobilized addressably in a re-usable oligonucleotide array. "Fl" stands for a fluorescent label; bars between nucleic acid sequences represent base-pairing; and wavy lines between the hexameric sequences and the substrates represent covalent linkages.

Fig. 7 is a schematic diagram illustrating a method of probing multiple (n) test RNAs using an oligonucleotide array (multi-wavelength array readout). Each test RNA is reverse transcribed with a primer distinctly labeled with a fluorophore. Bright squares in the array represent oligonucleotides that have ligated to fluorescently labeled DNA transcripts.

Fig. 8 is a schematic diagram illustrating an oligonucleotide array plate. The array plate includes an array of pins, each fitting into a well of a standard microtiter plate. The face of each pin contains an oligonucleotide array, which consists of addressably immobilized oligonucleotides (i.e., tags).

Fig. 9A is a schematic representation of the sequence of a test RNA (SEQ ID NO:44) to be structurally probed. DMS-methylated nucleotides are represented by (■).

Fig. 9B is a schematic representation showing the sequences (SEQ ID NOS:35, and 56-64) of the cDNA transcripts resulting from reverse transcription of the test RNA (SEQ ID NO:44) of Fig. 9A, and the top strand sequences (SEQ ID NOS:15 and 45-55) of their respective selection oligos.

Figs. 10A-10D are schematic diagrams illustrating four variations of a hybridization-based selection method in which the structure of a nucleic acid fragment is probed without transcription.

Fig. 11 is a schematic diagram and flow chart illustrating a new method for probing the ligand-binding state of a test RNA or DNA. The probing is performed with a microtiter filterplate system, which facilitates purification of the probed nucleic acids. "Stop" stands for a reagent that stops the probing reaction; and "PPT" stands for precipitation.

Fig. 12 is a diagram illustrating a drug screening protocol based on the method of Fig. 4.

Fig. 13 is a diagram illustrating a new ligation-dependent selection method for drug screening. The screening is performed with a microtiter filterplate system. "Stop" stands for a reagent that stops the probing reaction.

Detailed Description

The invention features methods of isolating or quantitating (i.e., determining the presence and/or amount) a target nucleic acid fragment contained in a mixture of nucleic acid fragments (including those from a biological sample such as blood or tissue samples). The

methods fall into two related, yet distinguishable categories: (1) hybridization-based selection methods; and (2) ligation-dependent selection methods. All of these methods rely on the use of a selection oligonucleotide ("oligo") that hybridizes to a specific sequence in the target fragment.

All oligos used herein can be either single-stranded or double-stranded. Optimal conditions for hybridization of oligos to target sequences can be readily determined by one of ordinary skill in the art. General guidance is provided in Sambrook et al. Molecular Cloning, A Laboratory Manual, 2nd Ed. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989), and Ausubel et al. Current Protocols in Molecular Biology, (Greene Publishing and Wiley-Interscience, New York, New York, 1993). An exemplary hybridization buffer contains 50 mM Tris (pH 7.5), 10-1000 mM KCl, 0-50% of formamide; and incubation can be performed at 4-60°C for 1-100 minutes. Stringency of a hybridization buffer can be controlled by varying, e.g., the concentration of salt, formamide, or both.

Oligos can be immobilized on a solid support prior, concurrent, or subsequent to hybridization. The solid support can be biological, non-biological, organic, inorganic, or a combination of any of these, existing as beads, strands, precipitates, gels, plates, etc. Examples of solid supports include, but are not limited to, agarose gel beads or columns, latex beads, plastic plates, and paramagnetic beads. Immobilization can be effected via, e.g., a non-covalent coupling. By way of example, biotinylated oligos can be immobilized onto a streptavidin or avidin-coated solid surface. Alternatively a covalent coupling can be employed. For instance, one can add a primary amino group to the oligo and covalently link it to a solid support derivatized

with a crosslinking agent such as an (N-hydroxysuccinamide-)-ester ("NHS-ester").

The new isolation/quantitation methods are applicable to nucleic acid structure probing, i.e., determination of the reactivity of a given nucleotide in a test nucleic acid toward a modifying agent. When structure probing is performed in the presence of a ligand (e.g., an inorganic or organic compound, another nucleic acid, a carbohydrate, or a polypeptide) or a potential ligand of a nucleic acid, additional information can be obtained about the biological functions of the nucleic acid and ligands thereof.

Hybridization-Based Selection Methods

The hybridization-based selection methods of the invention are used for isolating a target nucleic acid fragment from a mixture of nucleic acid fragments. The methods include two essential steps: substraction and selection.

During the subtraction step, a non-target fragment is removed by hybridizing it to an immobilized (including immobilizable) subtraction oligo that is complementary to a sequence present in the non-target fragment but absent in the target fragment. Of course, multiple non-target fragments can be removed simultaneously if they share that sequence. After hybridization, the aqueous phase, which contains the target and other remaining non-target fragments, is separated from the hybridized fragments, e.g., by centrifugation, filtration, magnetic field, or washing, depending on how the subtraction oligo is immobilized. The subtraction step is repeated, with different subtraction oligos, if necessary, until the target fragment has a unique sequence among the remaining fragments.

During the subsequent selection step, the aqueous phase recovered from the subtraction step is incubated with an immobilized selection oligo, which hybridizes to the unique sequence in the target fragment. The target fragment is then separated from the remaining fragments through separation of the solid and aqueous phases, as is performed in the subtraction step. Of course, if the target fragment is the only sequence left after the subtraction step(s), the selection step can be omitted.

Any given transcript of a transcription reaction, such as a primer extension reaction, can be isolated by the new hybridization-based selection methods. See, e.g., Example 1 and Figs. 2a-2d. The transcripts generated by such a reaction are progressively 3'-truncated fragments of the full-length transcript. Lengths of the transcripts depend on where the nucleic acid polymerase stalls during the transcription process. If the transcripts are labeled with a signal-producing agent, the amount of an isolated transcript can be quantified by measuring the level of the signal.

The hybridization-based selection methods require that the sequences involved, i.e., those intended to be hybridized by the selection or subtraction oligo, must be long enough (e.g., preferably no fewer than 6 nucleotides in length) and distinct enough to permit selective annealing reactions.

Nevertheless, these methods are highly flexible in other regards. For instance, the target fragment as well as some non-target fragments can first be isolated from the mixture by hybridizing them to an immobilized oligo specific for their common sequence. Subtraction and selection steps are then performed, as described above, to isolate the target from those non-target

fragments. Also, a target fragment in any given selection step need not be limited to a single fragment; multiple fragments having a common sequence can be isolated together as a target group by a selection oligo complementary to the common sequence.

Furthermore, one can detect the presence of, or quantify, multiple targets (related or unrelated) generated by different primer extensions in the same test tube. To do this, the targets are distinctly labelled by using distinctly labelled primers (e.g., different fluorescent tags) during primer extension. Therefore, the targets do not have to be separated from each other for quantitation.

EXAMPLE 1

Figs. 2a-2d illustrate a hybridization-based selection method for isolating a cDNA transcript of interest from a mixture of reverse transcription products. Among the four exemplary cDNA transcripts shown, one is a full-length cDNA transcript (SEQ ID NO:7), whereas the remaining three are truncated at A1408 (i.e., adenosine at position 1408; SEQ ID NO:8), A1492 (SEQ ID NO:3), and A1493 (SEQ ID NO:4), respectively. Since a ^{32}P end-labeled primer is used for the reverse transcription, each cDNA transcript is radioactive. The cDNA transcript truncated at A1408 is a target transcript (SEQ ID NO:8). The RNA template (SEQ ID NO:1) is degraded by RNase A after reverse transcription.

A subtraction step is subsequently performed to remove any transcript that is longer than the target. In this case, the longer transcript is the longest transcript of the reverse transcription. The subtraction oligo (5'-GGCGTCA-3') is biotinylated and immobilized onto a streptavidin-coated agarose bead (Sigma). After

annealing to the longest transcript, the bead is removed by various types of filtration devices, e.g., a cellulose acetate filter.

The supernatant is then placed in a streptavidin-coated microtiter plate onto which a biotinylated selection oligo, 5'-CACCTTCGGG-3' (SEQ ID NO:10) is immobilized. After the target anneals to the selection oligo, the plate is washed with, for example, a buffer containing 50 mM Tris (pH 7.5), 100 mM KCl, and 40% of formamide. This buffer is stringent enough to remove nonspecific hybridization, but preserves the specific hybridization between the selection oligo and the target.

Radioactive signals retained on the microtiter plate can be counted using a liquid scintillation counter, such as a PACKARD TOPCOUNT instrument, for quantifying the amount of the bound target.

All other transcripts can also be isolated and quantified in the same manner. To do so, a selection oligo that hybridizes to a 3' terminal sequence in the longest transcript is first employed to isolate and quantify that transcript. Subsequently, another selection oligo is used to isolate and quantify the new longest transcript. Thus, with a series of bead-bound oligos, one can scan, from the longest to the shortest, the entire family of the reverse transcription products based on their unique 3' ends.

Ligation-Dependent Selection Methods

In the new ligation-dependent selection methods, a target nucleic acid fragment with a single-stranded (ss) terminal sequence (ranging from, e.g., 2-10 nucleotides in length) can be separated from other nucleic acid fragments in a single selection step by a selection oligo.

The selection oligo is double-stranded (ds), with its upper or top strand comprising (i) a nonprotruding portion that complements to its lower or bottom strand, and (ii) a 5' or 3' protruding (i.e., single-stranded) portion that can anneal to the target's ss 3' or 5' terminal sequence. It is not necessary that the entire protruding portion of the oligo anneals perfectly to the target's ss terminal sequence: mismatches are permissible, so long as ligation between the oligo's lower strand and the target's ss terminal sequence is possible after annealing (see Example 2 and Fig. 3 described therein, *infra*). In general, the first five nucleotides in the oligo's protruding portion, i.e., the nucleotides that are immediately adjacent to the non-protruding portion, should be able to form Watson-Crick base-pairs (i.e., A-T, G-C, or A-U) with the 5 extreme terminal nucleotides in the target's ss terminal sequence. Of course, this number may vary, depending on various factors such as the sequences involved, the ligase used, and the ligation conditions. For instance, more than 5 perfect base-pairs may be needed at the site of ligation, when the ligation is performed at a higher temperature with use of a thermostable ligase such as *Taq* ligase.

Subsequent to annealing, the target's ss terminal sequence is ligated to the oligo's lower strand. Ligation can be performed, for example, in 20 μl of a ligation buffer (e.g., containing 50 mM Tris pH 7.2, 10 mM $MgCl_2$, 5 mM DTT, 1 mM ATP, 15% PEG 4000, 1 mM cobalt hexamine chloride, and 20 units of T4 DNA Ligase (New England Biolabs)) at 37°C for 1 hour. For increased ligation specificity, thermostable ligases such as *Taq* ligase can also be used. Ligated target fragments can be quantitated separately from the non-target fragments using techniques detailed later. Multiple fragments with

common ss terminal sequences can also be quantified as a group.

The characteristics of ligation alleviate the limitations inherent in the hybridization-based selection methods. This is because ligation effectively augments the specificity of nucleic acid hybridization with the high specificity of ligases for correct base pairing around the site of ligation. Thus, hybridization of the selection oligo to undesired nucleic acid fragments is harmless, as long as the terminal sequences of those fragments cannot anneal perfectly and ligate to the sticky end of the oligo. Also, since ligation occurs only at the ends of nucleic acids, it renders internal repeats of the 3'-terminal sequences substantially invisible. The high specificity of ligation reactions is evident in the experiments shown in Example 4, *infra*.

Due to this high specificity, the ligation-dependent selection methods can distinguish nucleic acid sequences that differ by as few as one nucleotide at their terminal regions. For instance, the only difference between a target and an undesired nucleic acid can be merely that the target has a different 3' terminal nucleotide, or is one nucleotide longer.

The following describes several variations of the new ligation-dependant selection methods for separating a ligated target nucleic acid fragment from unligated non-target fragments.

<u>The Selection Oligo's Bottom Strand is Immobilized</u>

If the selection oligo is immobilized onto a solid surface via its lower strand, non-target nucleic acid fragments (which are incapable of ligating to the oligo's bottom strand), and the oligo's upper strand are removed by a denaturing wash (see Example 2, *infra*). By way of example, the wash can be performed in a buffer

containing (i) 7 M urea and 0.5 X TBE/ (ii) 90% formamide, or (iii) 0.5M NaOH/3M NaCl at 55°C for 5 minutes. The target fragment remains attached to the oligo's immobilized lower strand via a covalent bond.

If the target has been labeled with a signal-producing agent (e.g., a radioactive or fluorescent agent, or an enzyme that catalyzes fluorescent reactions), signals retained on the immobilized lower strand after washing can be measured to quantify the amount of the recovered target.

The Selection Oligo's Upper Strand is Immobilized

If the selection oligo is immobilized onto a solid support via its upper strand, there are two ways of retaining the ligated target on the solid support.

The first way is to remove, by a semi-denaturing wash, nucleic acid fragments that hybridize to the selection oligo's protruding end but fail to link to the oligo covalently. This wash, however, preserves base-pairing between the upper and lower strands in at least a part of the selection oligo. Thus, the target fragment is selectively retained on the solid surface, since it is anchored to the surface through a covalent bond with the oligo's bottom strand, which remains base-paired with the upper strand. The stringency of the semi-denaturing wash depends on the stability of the base-paired region of the oligo itself, and can be determined empirically using standard methods. For instance, one can start by performing the wash in a buffer containing 0.5 X TBE and 1 M urea at 55°C for 5 minutes, and adjust the buffer composition (e.g., urea concentration), wash temperature, and/or wash time to achieve the best result.

To enhance the base-pairing strength in the double-stranded portion of the oligo, a G/C rich sequence

may be utilized, as the base-pair G/C is more stable than the basepair A/T or A/U.

Alternatively, if the double-stranded oligo is unimolecular (e.g., having a hairpin structure at its double-stranded end or being cross-linked at another part of the oligo), a regular denaturing wash will remove all of the unligated nucleic acid fragments, while retaining the ligated target. See, for instance, the following formulae:

```
GTGGGNNNNNN-3'---- Solid surface (SEQ ID NO:6)
(                                                   (1)
CACCC-5'


GTGGGGNNNNNN-3'---- Solid surface (SEQ ID NO:6)
(         ||||||                                    (2)
CACCCCnnnnnnnnn-5'              (SEQ ID NO:9)
```

Formula (1) shows a double-stranded oligo with a protruding 3' end NNNNNN, and the oligo's upper strand is immobilized via, e.g., its 3' end, onto a solid surface. Formula (2) shows that a target sequence 3'-nnnnnnnnn-5' hybridizes to the oligo's protruding end and is ligated to the oligo's lower strand. In both formulae, "(" denotes a covalent link (direct or indirect) between the oligo's two strands; and "|" denotes base-pairing. As illustrated in formula (2), the target sequence will be retained on the solid surface by the covalent link even after its base-pairing with the oligo is disrupted.

The Selection Oligo can be Immobilized in an Array

The selection oligos used for the ligation-dependent selection methods can be immobilized addressably, i.e., at a pre-defined location on a solid surface. Such a solid surface can have a high density of double-stranded selection oligos immobilized onto it, each oligo having a distinct single-stranded end and immobilized at a distinct, predefined location. Use of such high-density oligo arrays enables simultaneous isolation and quantification of multiple targets, each of which will be ligated to an appropriate oligo. The amount of each target can be determined by measuring the signal it produces at the pre-defined location where its corresponding oligo is immobilized.

Selection Oligos can be immobilized onto a solid surface addressably by methods known in the art, e.g., those disclosed in U.S. Patents Nos. 5,445,934, 5,510,270, 5,556,752, 5,143,854, and 5,412,087, and in PCT application WO 92/10092. As an example, the '854 patent discloses a light-directed method of forming oligos on solid surfaces or substrates. In this method, predefined regions of a surface are first activated by a light source, typically through a mask, much in the manner of photolithography techniques used in integrated circuit fabrication. The surface is subsequently contacted with a preselected nucleotide solution. Other regions of the surface remain inactive because they are blocked by the mask from illumination and remain chemically protected. By repeatedly activating different sets of predefined regions and contacting different nucleotide solutions with the surface, a diverse array of specific selection oligonucleotides is formed on the surface.

To generate a selection oligo array for the present ligation-mediated selection methods, one can, for

instance, first generate the top strands of these oligos, with the 3′ ends of the strands linked directly, or preferably through a spacer, to a solid surface. The spacer should have sufficient length (e.g., 6-50 atoms long) so as to allow the double-stranded oligos to interact freely with molecules exposed to the surface. The spacer can be, for example, aryl acetylene, ethylene oligos containing 2-14 monomer units, diamines, diacids, amino acids, polynucleotides, or combinations thereof (see, e.g., U.S. Patent No. 5,556,752).

After forming on the surface the oligos′ upper strands, each differing only in their 5′ or 3′ end, the common lower strand of the double-stranded oligos is loaded onto the support, resulting in an array of double-stranded oligos with common double-stranded portions and distinct 5′ or 3′ ss ends.

A mixture including target nucleic acid fragments can then be applied to the surface. The target fragments are allowed to ligate to their corresponding selection oligos via their ss terminal sequences, and are specifically retained on the surface by the above-described semi-denaturing wash (see also Example 5, *infra*).

The solid surface with the selection oligo array can be re-used after ridding the surface of non-covalently linked nucleic acids and reloading the lower strand of the oligos.

Alternatively, one can generate unimolecular double-stranded selection oligos (see, e.g., U.S. Patent No. 5,556,752). One exemplary unimolecular double-stranded selection oligo is shown in formula (3):

5′-CATGGCCGCACC-L-GGTGCGGCCATGNNNNNN-3′-Solid Surface (3)

(SEQ ID NOs:11 and 12, separated by -L-)

wherein L is optional, and represents a linker such as an alkylene group of from about 6 to about 24 carbons in length, or a polyethylene glycol group of from about 2 to about 24 ethylene glycol monomers in a linear configuration. The nucleotide sequences flanking L, which are G/C rich, will anneal to each other to form a double-stranded oligo with a hexameric protruding end. After ligation of a target fragment to the lower strand (i.e., the sequence left of "L" in formula (3); SEQ ID NO:11), a regular denaturing wash will remove all nucleic acids except the target fragment.

<u>The Target Nucleic Acid is Immobilized</u>

The above-described ligation-dependent selection methods can be modified to quantify a target nucleic acid fragment without separating it from other contaminant nucleic acid fragments in the mixture. To achieve this, the nucleic acid fragments in the mixture, instead of the selection oligo, are immobilized onto a solid surface. Moreover, the lower strand of the selection oligo is labelled with a signal-producing agent. After the ligation step, unligated oligos or oligo strands are removed by a denaturing wash. The level of signals retained on the solid surface positively correlates to the amount of the target present (see Example 3, *infra*).

If the upper, instead of lower, strand of the oligo is labelled with a signal-producing agent, a semi-denaturing wash can be performed to selectively retain double-stranded oligos that have ligated to an immobilized target; if the signal-producing upper strand is linked to the lower strand covalently, a regular wash can be performed instead.

EXAMPLE 2

Fig. 3 illustrates a ligation-dependent selection method for quantifying a cDNA transcript of interest. Purely for illustrative purposes, the nucleic acid fragments (SEQ ID NOS:2-5, 7, and 8) to be separated are depicted as products of a reverse transcription reaction, and the templates are modified RNA molecules. As described above, nucleic acid fragments to be separated and quantified in the present method need not be related in sequence.

The target transcript here is the longest one, which has a unique 3' terminal sequence of 5'-TGACGCC-3'. The selection oligo contains a highly G/C-rich hexameric clamp and a 3' sticky end with the sequence of 5'-GGCGTCA-3'. The bottom strand of the oligo is 5'-phosphorylated. After the target anneals to the sticky end, a ligase is added to seal the nick between the target and the oligo's lower strand. A subsequent strong denaturing wash (e.g., 8M urea and 0.1 X TBE at 37°C) results in dissociation of the top strand (SEQ ID NO:13) of the oligo, making retention of the 5'-end labeled cDNA on the filter completely dependent on the formation of a covalent bond between the bottom strand and the cDNA (SEQ ID NO:14).

EXAMPLE 3

Fig. 4 illustrates a new ligation-dependent selection method for simultaneous detection of multiple target nucleic acid fragments. This method does not require separating the target fragments from contaminant fragments. In this method, selection oligos (referred to below as "tags") for the target fragments are each labeled with a fluorophore (i.e., Fluor 1, 2, 3, or 4), which emits at a distinct wavelength (Molecular Probes, Eugene, Oregon). Preferably, the fluorophores selected

for use should have minimal overlap of emission maxima. Useful fluorophores include, but are not limited to, BODIPYFL (EM = 513nm), 4',5'-dichloro-2',7'-dimethoxyfluorescein (Em = 550nm), tetramethylrhodamine (Em = 580nm), and X-rhodamine (Em = 605nm).

A modified RNA is reverse transcribed with a 5'-biotinylated ("Bio") primer. The RNA component of the resulting RNA-DNA hybrids is then degraded by incubation at 55°C for 5-60 minutes with 10 units of RNase A. Selection oligos each labeled with Fluor 1, 2, 3, or 4 are then added along with T4 DNA ligase and a ligase buffer containing 50 mM Tris pH 7.5, 10 mM $MgCl_2$, 10 mM DTT, 1 mM ATP, 15% PEG 4000, and 1 mM cobalt hexamine chloride. Each selection oligo has a specific 3'-sticky end sequence complementary to the 3'-end of a target cDNA.

After ligation, streptavidin-coated agarose beads ("St") are added to immobilize the cDNAs, permitting removal of excess, unligated selection oligo tags as well as the oligo tags' top strands during the subsequent strong denaturing wash and filtration steps. Target cDNAs are thereby labeled and retained on the filter. Signals emitted by these cDNAs are quantitated with a microtiter plate reader equipped to measure fluorescence signals at multiple wavelengths.

EXAMPLE 4

Figs. 5a and 5b show the nucleotide sequences used in an experiment that examined the specificity of ligation. A double-stranded oligo ("tag") was created by annealing the single-stranded "top" (5'GGAGAACAGGAAGGGGCACCTT-3'; SEQ ID NO:15) and "bottom" (5'-CCCCTTCCTGTTCTCC-3'; SEQ ID NO:34) oligos. The tag contains a hexameric 3' protruding (or sticky) end in the top strand, and the hexameric sequence is complementary to the last 6

nucleotides at the 3' end of the XS oligo (5'-GCACAGCCTTGTTACGACTTCACCCGAAGGTG-3'; SEQ ID NO:35). XS was labelled with ^{32}P at its 5' end.

The ligation reactions contained 1 ng of XS oligo and a three-fold molar excess of tag in 20 μl of a buffer containing 50 mM Tris (pH 7.5), 10 mM $MgCl_2$, 10 mM DTT, 1 mM ATP, 15% PEG 4000, 1 mM cobalt hexamine chloride, and 1 Unit/μl T4 DNA ligase (New England Biolabs). The reactions proceeded for 1 hour at 37°C. The ligation products were precipitated with ethanol, resuspended in a loading buffer containing 8 M urea, and size-fractionated on a denaturing polyacrylamide gel. The gel was dried and exposed to X-ray film to generate an autoradiograph.

In a control reaction where the tag was omitted, XS migrated to the position indicated as "XS" in the gel, as shown in the autoradiograph. However, when the tag was added to the reaction, approximately 25% of the XS oligo shifted up to a position corresponding to the covalent addition of the bottom strand of the tag to the XS oligo.

Sequence sensitivity of ligation was investigated by mutating each of the six nucleotides at the tag's 3' protruding end. In total, 18 such mutants (SEQ ID NOS:16-33) were generated, three for each of the six positions (Fig. 5 top, where nucleotide mutations are framed). The data showed that the ligation reaction required correct base-pairing at positions 1 through 5, and that ligation between XS and a mutant oligo was reduced by at least one hundred fold. However, when the XS oligo was mutated so that perfect base pairing was obtained for those positions (SEQ ID NOS:35-43), the ligation efficiency was restored to a normal level. In contrast, no specificity was evident at position 6.

Finally, the data also show that ligation is highly sensitive to the length of the 3' terminal sequence (i.e., 5'-AAGGTG-3') of XS that complements to the protruding end of the tag. Ligation was undetectable when that terminal sequence was either lengthened by one base (i.e., 5'-AAGGTGT-3' in XS+1) or shortened by one base (i.e., 5'-AAGGT-3' in XS-1).

EXAMPLE 5

Fig. 6 illustrates a ligation-dependent selection method in which a selection oligo is immobilized via its upper strand onto a solid surface ("Substrate"). This method uses a two-dimensional array of pre-immobilized double-stranded oligos. Each oligo has a unique 3' hexameric single-stranded end for distributing labeled target nucleic acid fragments. The distributed target fragments are then quantitated by illuminating and optically scanning the surface of the array.

The nucleic acid fragments to be separated are obtained by reverse transcription of a modified RNA using a fluorescent- labeled primer. The RNA component of the resulting RNA-DNA hybrids is degraded as described above. The cDNA transcripts, T4 DNA ligase, and ligase buffer are then mixed and incubated with the oligo array. After an hour of incubation at 37°C, target cDNAs are specifically ligated to matching oligos in the two-dimensional array.

Next, a semi-denaturing wash is performed in a buffer containing 1M urea and 0.1 X TBE at 37°C. The washing conditions are adjusted to disrupt hexamer-cDNA base pairing, while leaving at least a portion of the oligo's own double-stranded region intact. This wash makes retention of the cDNAs on the array strictly dependent on the formation of a covalent bond between the

5'-end of the oligo's bottom strand and the 3'-end of the cDNA.

At this point, the array may be illuminated and optically scanned. After readout, the array is recycled by (i) subjecting it to a strongly denaturing wash which removes the ligated, bottom strand from the immobilized top strand, and (ii) reloading a new bottom strand to the array.

EXAMPLE 6

As described in Example 4, *supra*, T4 DNA ligase typically requires that the five 3' (or 5') terminal nucleotides of a DNA fragment base-pair with the 3' (or 5') sticky end of a double-stranded oligo at the ligation site. Based on this specificity, the procedure described in Example 5 will allow the identification of 1024 (i.e., 4^5) unique terminal sequences with the use of a single fluorescent label. This capacity can be increased, however, simply by employing multiple primers, each derivatized with a fluorophore emitting at a distinct wavelength, as illustrated in Fig. 7.

Fig. 7 shows a number of reverse transcription reactions, each using a different modified RNA template and a different primer distinctly labeled with a fluorophore. These reactions, however, are performed in a single container (e.g., well). In other words, the transcripts from these reactions are all mixed together. The combined cDNA products of the reactions are then simultaneously ligated to a random hexamer tag array, as in Fig. 6.

Transcripts from different reverse transcription may ligate to the same location in the array. Yet, these transcripts can be independently quantitated, if their fluorescent labels emit at different wavelengths. Thus, scanning the array

sequentially at each wavelength will allow approximately n x *1024* pentamers to be independently reported, where n is the number of primers. The main limitation on n will be the maximum number of existing differentiable fluorophores.

EXAMPLE 7

Fig. 8 illustrates a method of using the oligo array described in Examples 5 and 6 in conjunction with microtiter plates. In this method, the oligo's upper strands are first synthesized on the faces of pins that fit into standard microtiter plate wells. The synthesis utilizes conventional phosphoramidite DNA chemistry, with the addition of an initial light-directed coupling step to define the small geometries required.

Specifically, pin faces are first derivatized with a photolabile primary amino group. Subsequently, a photolithographic mask is used to expose a selected pin face area to light, thereby deprotecting, or otherwise activating, the amino group within that area. At this point, an activated phosphoramidite is coupled to the amino group, and standard DNA synthesis procedures are employed to couple and synthesize the DNA strand. This array of single-stranded DNAs is subsequently transformed to an array of double-stranded oligos with a 5' or 3' sticky ends by hybridizing the existing array with the ds oligos' common bottom strand.

Methods For Nucleic Acid Structure Probing

All of the above-described isolation/quantitation methods can be applied to probe nucleic acid structures by determining the reactivity of a given nucleotide in an RNA or DNA molecule toward a modifying reagent.

To achieve this, a plurality of the test nucleic acid molecules (fragments) are incubated with an appropriate modifying reagent to allow the modification to occur under conditions so that statistically no more than one modification occurs to each molecule. Subsequently, these test molecules are used as templates for transcription, which terminates at any appropriately modified nucleotide. If the nucleotide of interest has been modified by the reagent, a transcript ending at that nucleotide will be generated; as a result, the presence of such a transcript indicates that the nucleotide is reactive toward the modifying reagent (see, e.g., Fig. 1). That transcript can be the target for the isolation/quantitation methods of the invention.

Alternatively, test nucleic acid molecules that have been treated with a modifying reagent can be further treated with a cleaving agent that cleaves nucleic acids where modified. Cleaved fragments corresponding to a given modified nucleotide can be detected and quantitated by the isolation/quantitation methods of the invention.

As described above, the hybridization-based selection methods require that the sequences intended to be hybridized by the selection or subtraction oligo must be long enough (e.g., 6 or more nucleotides in length) to permit selective annealing. However, when these methods are used for structure probing, this requirement may not pose a significant problem. The base selectivity inherent in chemical probes tends to reduce the impact of this requirement, and the requirement applies only to adjacent identical bases. Nonetheless, even adjacent identical bases can be reported as a combined signal.

<u>Modification of Nucleic Acid Templates</u>

RNAs can be modified by chemical and/or enzymatic reagents so that they can be cleaved where

properly modified, or can serve as templates for reverse transcription which terminates at any appropriately modified nucleotide (Ehresmann et al., *Nucleic Acids Research*, 15: 9109-9128, 1987; Stern et al., *Methods in Enzymology*, 164: 481-489, 1988). For example, N1 of A and N3 of C can be modified by DMS; N1 of G and N3 of U by 1-cyclohexyl-3-(2-morpholinoethyl)-carbodiimide metho-p-toluene sulfonate (CMCT); and N1 and N2 of G by β-ethoxy-α-ketobutyraldehyde (kethoxal). These modifications can all stall RT. Modification of phosphates of any nucleotides by ethylnitrosourea, modification of N7 of A by DEPC, and modification of N7 of G by DMS, on the other hand, are not sufficient to block DNA elongation by RT, but further chemical treatment, such as reduction with sodium borohydride and a further incubation with a base (aniline for RNA, and pyridine for DNA), can be employed to create strand breaks. Additional modifying reagents include, but are not limited to, bisulfite and methidiumpropyl-EDTA. Enzymatic reagents, such as nuclease S1, *Neurospora crassa* nuclease, and RNases T1, U2, Cl3, T2, and V1, can also be used to cleave RNAs at various specific sites (see, e.g., Ehresmann, *supra*).

Reagents for generating DNA modifications include, but are not limited to, DMS, piperidine, hydrazine, and $KMnO_4$ (Sasse-Dwight et al., *Methods in Enzymology*, 208: 146-168, 1991; Sambrook et al., *supra*). DMS modifies, among other sites, N7 of G. When treated appropriately with piperidine, the DNA strand becomes susceptible to cleavage at modified Gs. $KMnO_4$, on the other hand, modifies primarily T and to a lesser extent C. Treatment of the modified DNA with alkali results in the conversion of the T residue to urea, which is unable to be copied by a DNA polymerase. Base modification or removal by hydrazine and piperidine are described in

detail by Sambrook et al. Aside from chemical compounds, enzymes such as DNase I and micrococcal nuclease can also be used to generate DNA strand breaks.

DNA or RNA targets can be modified *in vitro* as well as *in vivo*. *In vivo* DNA or RNA targets include, but are not limited to, bacterial ribosomal RNAs, viral RNAs such as HIV RRE and TAR in human cells, cellular RNAs such as telomerase RNA, and viral or cellular transcriptional cis elements. If the chemical or enzymatic modification is to be performed *in vivo*, the chemical or enzyme of choice should be able to penetrate the cell, or it can be delivered to the cell by well known techniques such as liposome fusion, erythrocyte ghosts, or microsphere (microparticles; see, e.g., U.S. Patent No. 4,789,734).

Transcription of Nucleic Acid Templates

Transcription procedures using DNA or RNA as a template are well known in the art (see, e.g, Sambrook et al., *supra*). Transcription of RNA templates utilizes reverse transcriptases, which extend a DNA primer from a position 3' of the RNA region to be monitored. To facilitate subsequent hybridization between the cDNA transcripts and their corresponding subtraction or selection oligos, the RNA templates are preferably degraded after transcription. Degradation can be achieved by, for example, incubation at 55°C for 5-60 minutes in the presence of 10 units of RNase A.

Transcription of DNA can be accomplished by DNA-dependent DNA polymerases such as Klenow fragment and *Taq* polymerase in primer extension reactions. To increase the yield of free transcripts (i.e., transcripts that do not remain hybridized to their templates) in primer extension reactions, DNA templates can be repeatedly separated from their extension products by

heat-denaturing, and re-annealed to free primers for another round of primer extension, much in the manner of a polymerase chain reaction; a thermostable polymerase such as *Taq* (Hoffmann La-Roche), *Pfu* (Strategene), or VENT (New England Biolabs) polymerase can be used for this purpose.

Alternatively, DNA-dependent RNA polymerases such as T7 RNA polymerase are used. These RNA polymerases require promoters, but not primers, for transcription. The promoters, which are double-stranded (e.g., 10-20 base-pairs), can be attached to a template region by standard cloning techniques. For instance, a partially double-stranded DNA construct can be made which includes a shorter top DNA strand containing a promoter sequence, and a longer bottom strand containing a complementary promoter sequence and a template sequence.

Signal-producing reagents such as radioactive isotopes (e.g., ^{32}P and ^{35}S), fluorescent reagents (e.g., fluorescein, phycoerythrin, Texas Red, or Allophycocyanin) and enzymes catalyzing fluorescent reactions (e.g., horseradish peroxidase and alkaline phosphatase) can be incorporated into transcripts in multiple ways. For instance, the reagents can be used to label the oligonucleotide primers, or to label certain deoxyribonucleotide triphosphates (e.g., ^{32}P-dCTP or ^{35}S-dATP). Methods for detecting radioactive or fluorescent signals are well known in the art.

EXAMPLE 8

Figs. 9A and 9B illustrate an experiment that demonstrates the high specificity of ligation in a new structure probing method. The results shown in Example 4 were extended in the experiment shown in this example.

In this example, cDNAs generated by reverse transcription of a DMS-modified RNA (SEQ ID NO:44) were

ligated to specific selection oligo tags. Fig. 9A shows the secondary structure of the RNA with sites of DMS methylation indicated by filled squares (■). The modified RNA was reverse transcribed, from a site 3' to the structure shown, with a 5'-end labeled primer using standard reaction conditions. cDNA products (SED ID NOs:35 and 56-64) are listed in Fig. 9B. The RNA template was degraded after transcription by incubation at 55°C for 15 minutes with 10 units of RNase A.

Selection oligos (10 pmoles) for each of the cDNAs and 20 units of T4 DNA ligase (NEB) were added directly to the RT reaction mix (10 μl) after the reaction was stopped. The final ligation reaction mix was 20 μl in volume and was in a 1X ligation buffer (100 mM Tris pH 7.6, 10 mM $MgCl_2$, 10 mM DTT, 1 mM ATP, 15% PEG 4000, 1 mM cobalt hexamine chloride). The selection oligos' bottom strand (SEQ ID NO:34) was identical to that shown in Fig. 5A, and the oligos' top strand sequences (SEQ ID NOs:15 and 45-55) are shown in Fig. 9B. The negative controls of this experiment were (i) reverse transcription of unmodified RNA (K), and (ii) ligation of the modified RNA without any selection oligo.

After 1 hour of incubation at 37°C, ligation products were precipitated by ethanol, resuspended in a loading buffer containing 8 M urea, and size-fractionated on denaturing polyacrylamide gel. As shown in an autoradiograph of the gel, the cDNAs corresponding to modification of C1407, A1408, A1492, and A1493 were approximately quantitatively ligated to their respective tags, as the bands corresponding to their unligated cDNA's were strongly depleted. Meanwhile, new bands appeared at positions corresponding to the addition of the 16 nucleotide bottom strand of the tag to its respective cDNA target. This addition appeared highly specific, since only cDNAs targeted by the tag were

depleted and shifted. Ligation of tags to cDNAs A1499-A1503 was also strong and specific, but depletion was less evident, presumably because these cDNAs were present at much higher stoichiometries. The remaining three cDNAs, i.e., A1398, C1402, and C1403, were present at much lower stoichiometries, and their ligated products were not visible on the gel. However, specific depletion of A1398 was evident.

EXAMPLE 9

Figs. 10A-10D show four modes or variations of a new hybridization-based selection method in which a DNA fragment is structurally probed without transcription. The DNA, which is labeled at its 3′ terminus, is cleaved with a backbone-cleaving enzymatic probe DNase I. Figs. 10A-10D are distinguished by where the initial subtraction or selection is made with respect to the targeted position for fragmentation within the modified DNA (the targeted position is indicated with the word "Quantitate").

In Fig. 10A, the 3′ cleavage product corresponding to the targeted position is quantitated. The initial step immobilizes all 5′ cleavage products on bead 1 by a first subtraction oligo, leaving all 3′ cleavage products in the mobile phase. 3′ cleavage products longer than the target are then removed from the mobile phase by their selective immobilization on bead 2, which contains a second subtraction oligo. At this stage, the target 3′ cleavage product possesses a unique 5′ terminal sequence which can be selectively hybridized with a selection oligo immobilized on bead 3. The amount of the target product recovered on bead 3 can then be quantitated.

In the variation shown in Fig. 10B, the initial position of subtraction is shifted to a region

immediately 5' to the target cleavage product. This results in the target product being the longest labeled fragment in the mobile phase, eliminating the need for a second subtraction step. A selection oligo immobilized on bead 2 is then used to select the target for quantitation.

In Fig. 10C, the initial subtraction position is shifted again to now overlap the 5' end of the target product. The targeted product, as well as other products immobilized on bead 1, is separated from the mobile phase and retained. The target product is now unique, since it is the only immobilized fragment without any extra 5' sequences. The other fragments are then subtracted away from the target via hybridization between their extra 5' sequences and a subtraction oligo on bead 2, which can be physically separated from bead 1. Beads 1 and 2 can, for instance, be an agarose bead, and a paramagnetic bead, respectively, so that application of a magnetic field can physically restrain bead 2 while bead 1 is aspirated out.

The variation shown in Fig. 10D is similar to that shown in Fig. 10C except that the initial subtraction position with bead 1 is 3'-shifted. This results in the immobilization of all 3' cleavage products. In the second step, products longer than the target are removed with bead 2 as illustrated in Fig. 10C. Finally, the target product is removed by bead 3 from the shorter product, which is linked only to bead 1.

Similar steps can be performed to analyze a modified nucleic acid fragment that is labeled at its 5' instead of 3' end.

Uses Of The Structure Probing Methods

The structure probing methods of the invention can in turn be applied to determine whether a compound

can alter the reactivity of a given nucleotide in an RNA or DNA molecule (fragment) toward a modifying agent. For instance, if a compound blocks the modification of an otherwise modifiable nucleotide, primer extension products that terminate at that nucleotide will be absent or decrease in amount. Conversely, when a compound enhances the reactivity of a nucleotide toward a modifying agent, the amount of primer extension products terminating at that nucleotide will increase. Thus, the new structure probing methods can be used in the following exemplary contexts: (i) screening for nucleic acid-binding compounds, (ii) aptamer screening for non-nucleic acid targets, (iii) characterization of Qualitative Structure-Activity Relationship ("QSAR"), and (iv) regulatory network profiling.

<u>Screening for Nucleic Acid-Binding Compounds</u>

The structure probing methods can be used to screen for nucleic acid-binding compounds (e.g., synthetic organic compounds, nucleic acids, polypeptides, and carbohydrates) (see Examples 10 and 11, *infra*), and to study the interaction between a nucleic acid molecule and its ligand.

Structure probing provides highly detailed and direct information about nucleic acid reactivities and ligand binding state. Furthermore, its capacity to report on the order of, for example, n x 1024 probing signals per well (with array readout techniques) means that very long nucleic acid sequences, or multiple short sequences, can be used as binding targets with no diminution in the quality of the information generated. Both of these characteristics are probably essential for successful screening for sequence-specific nucleic acid binding compounds.

For example, a highly specific compound in a library might be expected to recognize a nucleic acid target specified by 6 nucleotides. Such a compound will, by definition, bind tightly to only one, or a few, elements out of $4^6 = 4096$ possible ones (each element containing 6 or more bases). Since one cannot predict in advance which elements out of the 4,096 possible ones will be bound with high affinity, it will be necessary to examine a large fraction of them. This becomes feasible with the new high-throughput structure probing methods which can generate up to approximately 10,000 (i.e., 10 x 1024, when 10 wavelengths are used) structure probing signals per well, since 10,000 signals can report on, for example, 3,300 elements at 3 signals/element.

<u>Aptamer Screening for Non-Nucleic Acid Targets</u>

The new methods can be used to screen large numbers of compounds (e.g., nucleic acids, proteins, polysaccharides, and small organic compounds) for their ability to bind to non-nucleic acid targets. Appropriate aptamers are used in lieu of the non-nucleic acid targets in these screenings. Aptamers are RNA molecules selected from large random libraries on the basis of their ability to bind specific molecular targets (e.g., proteins, lipids, carbohydrates, steroids, nucleic acids, etc.). Aptamers are "molecular mimics" in the sense that they effectively imitate the binding characteristics of other molecules. For example, if a receptor R binds ligand L, then an aptamer selected on the basis of its interaction with L might be expected to be mimicking the region of R that binds L. This is particularly true if one uses R to elute potential aptamers from L during the selection procedure. For this reason, screenings of an R-mimicking aptamer against compound libraries can be expected to identify compounds that will also bind R. In this way,

aptamer screening effectively expands the range of targets amenable to structure probing analysis to all those for which an aptamer exists.

RNAs capable of binding to a specific target can be obtained by, for example, the Systemic Evolution of Ligands by Exponential Enrichment ("SELEX") technique described in U.S. Patent Nos. 5,475,096, 5,595,877, and 5,270,163, and Gold et al. *Annu. Rev. Biochem.*, 64: 763-797, 1995. For instance, the '163 patent describes a method in which a candidate mixture of single-stranded nucleic acids having regions of randomized sequence is contacted with a target compound; those nucleic acids having a higher affinity to the target are partitioned from the remainder of the candidate mixture; and the partitioned nucleic acids are then amplified to identify a target-binding nucleic acid.

Characterization of QSAR

In addition, the new methods can be used to characterize combinatorial libraries of synthetic compounds whose functions are unknown. To achieve this, the synthetic compounds are assessed for their ability to bind a panel of nucleic acid molecules, i.e., their ability to alter reactivities of certain nucleotides in the nucleic acid molecules toward modifying agents. A synthetic compound's nucleic acid-interacting profile is then compared with the corresponding profiles of known compounds. A similarity in the profile between the synthetic compound and a known compound indicates a similarity in function. Thus, there are two steps to the process: (1) screening known compounds to generate their nucleic acid-binding profiles; and (2) screening compound libraries to find compounds with similar profiles.

With the new methods, up to approximately 10,000 footprinting signals can be generated for a single

compound (or group of compounds) in a single microtiter plate well. These signals constitute a nucleic acid-interacting profile for that compound, locating it within a 10,000-dimensional structure probing space. This information may be combined with more traditional molecular indices to extract structure-activity relationships with well known statistical methods.

An exemplary nucleic acid that can be used to generate such profiles is 16S ribosomal RNA ("rRNA"). A number of other RNAs, proteins, and drugs (e.g., spectinomycin, tetracycline, streptomycin and neomycin) have been found to interact 16S rRNA. If an unknown compound has a substantially similar 16S rRNA-interacting profile as, say, neomycin, this unknown compound is then a likely substitute of neomycin.

To enhance the diversity of nucleic acids' three-dimensional configurations, and thereby their usefulness, one can either covalently derivatize, or noncovalently complex, more diverse molecules to them. For example, a nucleic acid target can be covalently derivatized with a protein, lipid, steroid, or carbohydrate moiety to increase the likelihood of generating a footprinting signal. One can also use nucleic acids with bound ligands (e.g., the decoding region of 16S rRNA complexed with neomycin) to generate compound profiles, to probe for alterations of those binding interactions by test compounds. Similarly, aptamers can be used to increase the effective diversity of nucleic acid targets in such applications.

In addition to synthetic combinatorial libraries, natural product libraries, such as extracts from various streptomyces fungi, can also be characterized as described herein.

Regulatory Network Profiling

At any given time point, a determination of the occupancy or binding status of *cis* regulatory elements (DNA and RNA) in a cell or virus constitutes a snapshot of the state of its genetic regulatory network. Monitoring the network state can provide detailed information about how the cell or virus responds to external stimuli (for example, drugs, or growth factors), genetic alterations (such as oncogenic transformation), or progression through the cell cycle. Monitoring can be accomplished using the new structure probing methods to screen various cellular extracts (e.g., mammalian cell extracts) against such a collection of *cis*-elements. For instance, to determine whether a given promoter is bound by proteins following a drug treatment of cells, one can determine whether the drug treatment leads to alteration of reactivity of selected nucleotides in the promoter toward a modifying reagent such as DNAse I.

EXAMPLE 10

Modified (or "probed") nucleic acids, or their corresponding cDNAs, can be generated using well-known laboratory techniques that are adapted to a microtiter plate format. Fig. 11 illustrates one such adaptation for studying RNA-ligand interaction.

A nucleic acid-ligand complex is first formed under appropriate buffer conditions in steps 1 and 2, and a probing reagent (e.g., DMS) is added in step 3. Specifically, an aliquot of approximately one picomole of a ligand, such as the aminoglycoside antibiotic neomycin, is added to the well (step 1). Then 100 μl of a binding mix containing 10-100 ng of an RNA 50mer target in 80 mM K-Hepes (pH 7.9), 20 mM $MgCl_2$, 100 mM KCl, 5% PEG 8000 is added (step 2). After a short equilibration period of approximately 5 minutes at 37°C, 0.5 μl of DMS is added.

The incubation proceeds for 5 more minutes at 37°C (step 3).

The probing incubation is stopped by addition of 10 μl 3 M NaOAc and 300 μl ethanol. The ethanol-precipitated RNA is then filtered, and retained on the filter membrane (step 5). The RNA is then either resuspended in a buffer (step 6a), or transcribed into cDNA fragments with a polymerase (step 6b).

The filtration step (and thus the plate) can be eliminated if the nucleic acid to be probed is immobilized in the well through the use of, for example, biotin-streptavidin coupling. In that case, the probing reaction can be stopped with a simple wash step.

EXAMPLE 11

Fig. 12 outlines a protocol by which drug compounds are screened for their ability to bind and protect a particular nucleotide in the RNA target from DMS modification. The protocol is based on the hybridization-based selection method illustrated in Fig. 2, and is as follows.

First, in the probing ("footprint") reaction, the RNA of interest is biotinylated and immobilized onto the wells of a microtiter plate at 10-100 ng RNA/well. Candidate drug compounds are then added to the wells in 15-100 μl of a buffer containing 80 mM K-Hepes (pH 7.9) and 50 mM KCl to create a binding mix (steps 1 and 2). 1 μl DMS is then added (step 3). After DMS modification, the RNA is purified by washing the plate with an appropriate solution with TE (step 4).

Next, the RNA is reverse transcribed. A ^{32}P 5'-end-labeled primer is annealed to the RNA and extended with RT by the addition of the required components of the RT reaction (step 5). After reverse transcription,

RNases H, A, or T1 are added to the RT reaction mixture to degrade the RNA template (step 6).

Then, in the subtraction/selection step, a subtraction oligo immobilized on agarose beads is added to the reaction mix to allow hybridization between the oligo and its target sequence (step 6). After hybridization, the beads are removed by filtration. The supernatant is recovered and transferred to a well in another microtiter plate coated with streptavidin (step 7).

A selection oligo is then added to the well (step 8). This oligo is biotinylated so that, once annealed, the transcript of interest (the synthesis of which terminates at the nucleotide of interest) is effectively immobilized in the well. All other (shorter) transcripts are washed away with TE or other appropriate buffers (step 9). The ^{32}P signals retained in the well are then counted using a PACKARD TOPCOUNT instrument (step 10). An absence of signals, or a decrease in signals as compared to a control sample which does not undergo the drug addition step, indicates that the drug compound is capable of blocking, or partly blocking, the chemical modification of the nucleotide of interest.

EXAMPLE 12

Fig. 13 illustrates a drug screening protocol based on the ligation-mediated selection method with differential labeling shown in Fig. 4. In this protocol, drug compounds are screened for their ability to bind to an RNA.

Drug compounds are first aliquoted into the wells of a microtiter filter plate (step 1). Target RNA(s) in 100 μl of a binding buffer (80 mM K-Hepes (pH 7.9), 1 mM $MgCl_2$, 100 mM KCl) are added to each well at a concentration of 0.5 μM (step 2). Incubation is

performed at room temperature for 5 minutes. 1-5 μl DMS is then added to each well, and incubation is continued for 5 additional minutes at room temperature (step 3). The reaction is stopped by incubation with 10 μl of 0.3 M NaOAc and 300 μl ethanol for 5 minutes at room temperature (step 4). Filtration of the precipitated RNA deposits the DMS-treated RNA onto a filter (step 5). The filtrate is discarded.

A RT mix containing 7.5 ng of a biotinylated primer is then added to re-suspend the RNA deposit on the filter (step 6). The RNA is incubated with the RT mix for 5-30 minutes at 37°C for reverse transcription. Subsequent to reverse transcription, RNase A is added at 2-10 units per μl in 50 mM Tris pH 7.5 to degrade the RNA templates. The incubation proceeds for 15 minutes at 55°C (step 7). Next, a mixture containing fluorescence-labeled DNA selection oligos, ATP, and a DNA ligase is also added, and ligation proceeds at 37°C for 1 hour (step 7). The selection oligos are added at approximately a 3-fold molar excess (e.g., approximately 10 picomoles each) to their respective cDNAs.

Subsequent to annealing and ligation, 10 μl of streptavidin-coated agarose beads are added in a denaturing buffer containing 6 M urea in 0.5 X TBE to immobilize the cDNAs and remove upper strands of the selection oligos (step 8). The denaturing buffer used in this step should be designed not to disrupt the streptavidin-biotin interaction. Filtration is used to remove everything except the immobilized cDNAs, which are then quantitated at different wavelengths with a fluorescence spectrometer (steps 9 and 10). The peaks 1-4 in the graph at the lower right corner of the figure correspond to the different fluorescent labels of the four selection oligos.

As used in this Example, top strands of the selection oligos are gel-purified synthetic DNAs. The bottom strands are gel-purified and 5'-phosphorylated with a polynucleotide kinase. In addition, the bottom strands carry fluorophore moieties, which can be attached, e.g., at the 3'-end by using amino-modified oligo. The amino group forms a covalent bond with a fluorophore molecule (available from Molecular Probes) carrying an NHS-ester group. Useful fluorophores include BODIPY FL, 4',5'-dichloro-2',7'-dimethoxyfluorescein, tetramethylrhodamine, and X-rhodamine, all of which are NHS-esters that will covalently crosslink to amino-modified oligos.

Other Embodiments

It is to be understood that while the invention has been described in conjunction with the detailed description thereof, the foregoing description is intended to illustrate but not limit the scope of the invention, which is defined by the scope of the appended claims.

Other aspects, advantages, and modifications are within the scope of the following claims.

What is claimed is:

1. A method of isolating a target nucleic acid fragment from a mixture of nucleic acid fragments, the method comprising:

removing a non-target fragment by hybridizing it to an immobilized subtraction oligonucleotide that is complementary to a sequence present in the non-target fragment but absent in the target fragment;

repeating the removing step to remove additional non-target fragments, until a known sequence in the target fragment becomes unique among the remaining fragments; and

selecting the target fragment by hybridizing it to an immobilized selection oligonucleotide that is complementary to the unique sequence.

2. The method of claim 1, wherein the nucleic acid fragments, including the target, are products of a transcription reaction.

3. The method of claim 2, wherein each of the nucleic acid fragments, including the target, comprises a signal-producing agent.

4. A method of determining whether a given nucleotide in a nucleic acid is modified by a modifying agent, the method comprising:

incubating the nucleic acid with the modifying agent under conditions that allow modification of the nucleic acid;

generating nucleic acid fragments from the incubated nucleic acid, wherein modification of the given nucleotide results in a target nucleic acid fragment whose terminus corresponds to the given nucleotide; and

isolating the target fragment, if present, using the method of claim 1, wherein the presence of the target fragment indicates that the given nucleotide is modified by the modifying agent.

5. The method of claim 4, wherein each of the nucleic acid fragments comprises a signal-producing agent, and the presence of the target is indicated by the signal detected after the isolating step.

6. The method of claim 4, wherein the generating step comprises contacting the incubated nucleic acid with a cleaving agent that cleaves nucleic acids only at a modified nucleotide.

7. The method of claim 4, wherein the generating step comprises transcribing the incubated nucleic acid, the transcription terminating at a modified nucleotide.

8. The method of claim 7, wherein the nucleic acid is an RNA, and the method further comprises degrading the nucleic acid with an RNase immediately following the generating step.

9. The method of claim 7, wherein the nucleic acid is a DNA, and the generating step further comprises, after the transcribing step, denaturing a nucleic acid duplex formed by the incubated nucleic acid and its transcript, and annealing the incubated nucleic acid with a primer for another round of transcription.

10. A method of determining whether a compound alters the reactivity of a given nucleotide in a test

nucleic acid toward a modifying agent, the method comprising:

incubating the test nucleic acid with the compound; and

determining whether the given nucleotide is modified by the modifying agent, using the method of claim 4;

wherein a change in the modifiability of the given nucleotide following treatment with the compound indicates that the compound alters the reactivity of the given nucleotide toward the modifying agent.

11. A method of quantitating a target nucleic acid fragment from a mixture of nucleic acid fragments, the target fragment including a unique, single-stranded terminal sequence, the method comprising:

obtaining a double-stranded oligonucleotide ("oligo"), wherein the oligo comprises (i) a first strand comprising a protruding portion that complements the unique terminal sequence, and a non-protruding portion; and (ii) a second strand that is complementary to the non-protruding portion;

incubating the mixture with the oligo to allow hybridization between the protruding portion and the unique terminal sequence;

ligating the unique terminal sequence of the target with the second strand of the oligo;

removing nucleic acid fragments that are not ligated to the second strand, thereby separating the target fragment from non-target fragments; and

measuring the amount of the target fragment to quantitate the target fragment.

12. The method of claim 11, wherein the target comprises a signal-producing agent,

the oligo is immobilized onto a solid surface via one or both of its strands; and

the measuring step detects the signal retained on the solid surface, wherein a level of the signal corresponds to the amount of the target present in the mixture.

13. The method of claim 11, wherein

the nucleic acid fragments, including the target, are immobilized onto a solid surface;

the oligo is labeled with a signal-producing agent; and

the measuring step detects the signal retained on the solid surface, wherein a level of the signal corresponds to the amount of the target present in the mixture.

14. The method of claim 11, wherein the first strand and the second strand of the oligo are covalently linked.

15. The method of claim 12, wherein the first strand is immobilized, and the removing step comprises incubating the mixture with a buffer that denatures base-pairing between the protruding portion and any other nucleotide sequence, but that does not denature base-pairing between the non-protruding portion and the second strand in a region of the oligo.

16. The method of claim 15, wherein the solid surface comprises a plurality of double-stranded nucleotide sequences, each of which, including the oligo, is immobilized at a distinct and pre-determined location on the solid surface.

17. The method of claim 15, wherein the region of the oligo is G/C rich.

18. The method of claim 13, wherein the first strand is labeled, and the removing step comprises incubating the mixture with a buffer that denatures base-pairing between the protruding portion and any other nucleotide sequence, but that does not denature base-pairing between the non-protruding portion and the second strand in a region of the oligo.

19. The method of claim 18, wherein the region of the oligo is G/C rich.

20. A method of determining whether a given nucleotide in a nucleic acid is modified by a modifying agent, the method comprising:

incubating the nucleic acid with the modifying agent under conditions that allow modification of the nucleic acid to occur;

generating nucleic acid fragments from the incubated nucleic acid, wherein modification of the given nucleotide results in a target nucleic acid fragment whose terminus corresponds to the given nucleotide; and

quantitating the target, using the method of claim 11, wherein the presence of the target indicates that the given nucleotide is modified by the modifying agent.

21. The method of claim 20, wherein the generating step comprises contacting the incubated nucleic acid with a cleaving agent that cleaves nucleic acids only at a modified nucleotide.

22. The method of claim 20, wherein the generating step comprises transcribing the incubated nucleic acid, the transcription terminating at a modified nucleotide.

23. The method of claim 22, wherein the nucleic acid is an RNA, and the method further comprises degrading the nucleic acid with an RNase immediately following the generating step.

24. The method of claim 22, wherein the nucleic acid is a DNA, and the generating step further comprises, after the transcribing step, denaturing a nucleic acid duplex formed by the incubated nucleic acid and its transcript, and annealing the incubated nucleic acid with a primer for another round of transcription.

25. A method of determining whether a compound alters the reactivity of a given nucleotide in a test nucleic acid toward a modifying agent, the method comprising:

incubating the test nucleic acid with the compound; and

determining whether the given nucleotide is modified by the modifying agent, using the method of claim 20;

wherein a change in the modifiability of the given nucleotide following treatment with the compound indicates that the compound alters the reactivity of the given nucleotide toward the modifying agent.

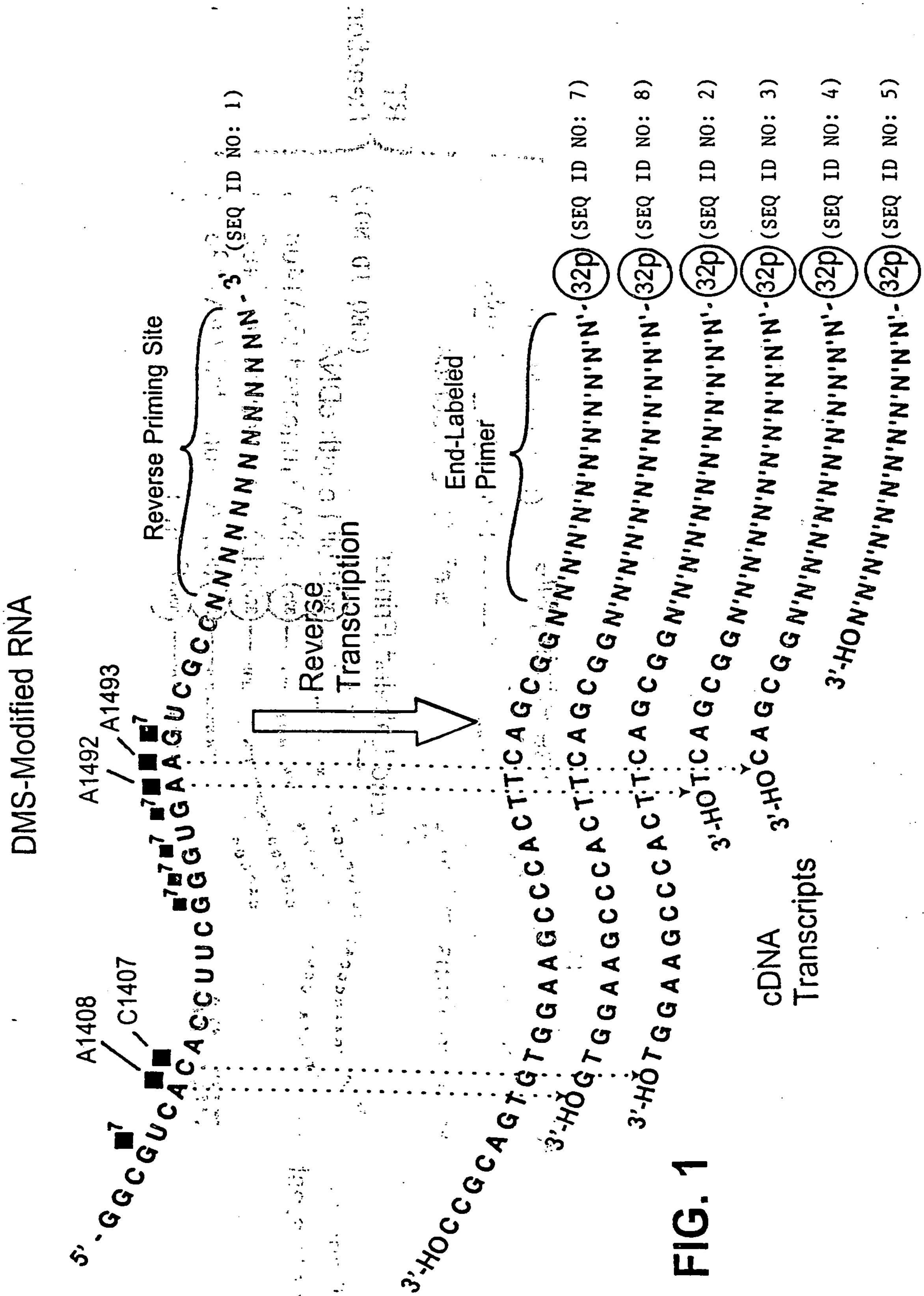

DMS-Modified RNA
A1408
C1407
A1492
A1493
Reverse Priming Site
(SEQ ID NO: 1)
Reverse Transcription
End-Labeled Primer
(SEQ ID NO: 7)
(SEQ ID NO: 8)
(SEQ ID NO: 2)
(SEQ ID NO: 3)
(SEQ ID NO: 4)
(SEQ ID NO: 5)
cDNA Transcripts

FIG. 1

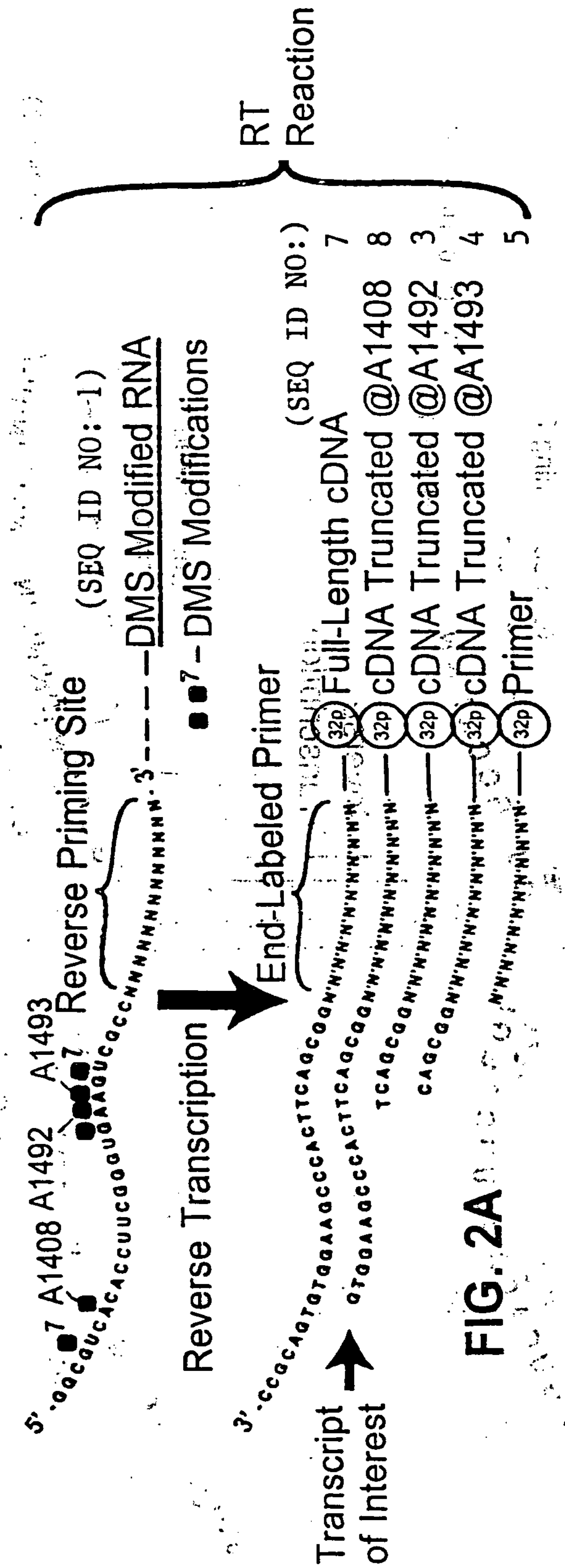

5'-GGCGUCACACCUUUCGGGUGAAGUCGCCNNNNNNNNNNNNN-3'
A1408 A1492 A1493
Reverse Priming Site
(SEQ ID NO:-1)
DMS Modified RNA
7 - DMS Modifications
Reverse Transcription
3'-CCGCAGTGTGGAAAGCCCACTTCAGCGGN'N'N'N'N'N'N'N'N'N'N'N'N'
End-Labeled Primer
RT Reaction
(SEQ ID NO:)
32p Full-Length cDNA 7
32p cDNA Truncated @A1408 8
32p cDNA Truncated @A1492 3
32p cDNA Truncated @A1493 4
32p Primer 5
Transcript of Interest

FIG. 2A

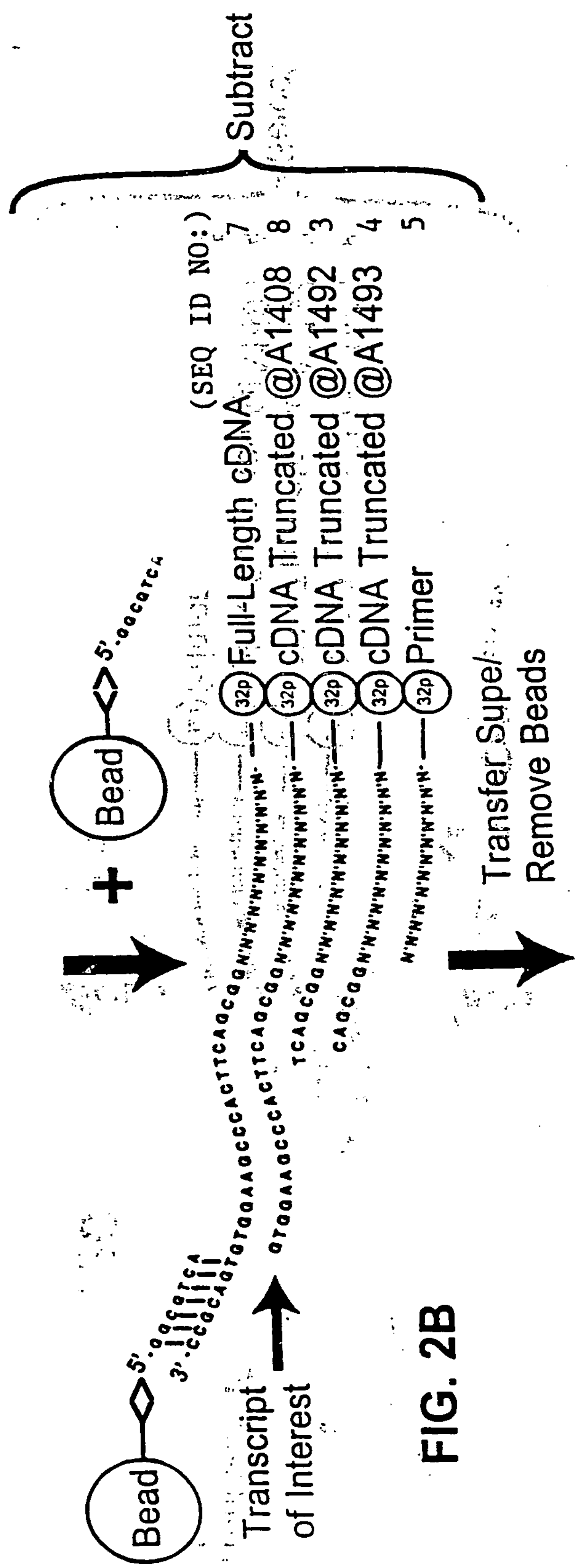

Bead
Transcript of Interest
+
Bead
Full-Length cDNA
cDNA Truncated @A1408
cDNA Truncated @A1492
cDNA Truncated @A1493
Primer
(SEQ ID NO:)
7
8
3
4
5
Subtract
Transfer Supe/
Remove Beads

FIG. 2B

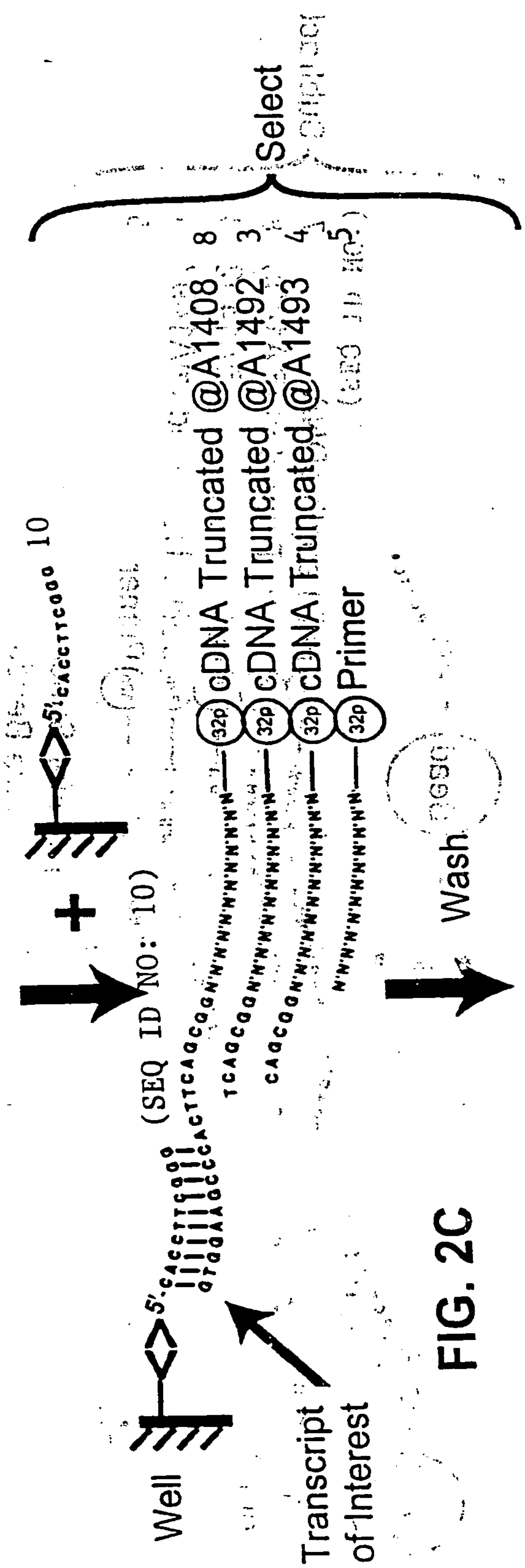

Well
Transcript of Interest
(SEQ ID NO: 10)
10
Wash
cDNA Truncated @A1408 8
cDNA Truncated @A1492 3
cDNA Truncated @A1493 4
Primer
Select
32p

FIG. 2C

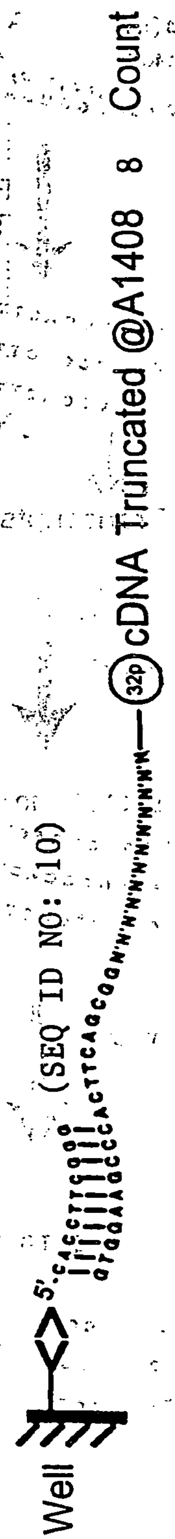


FIG. 2D

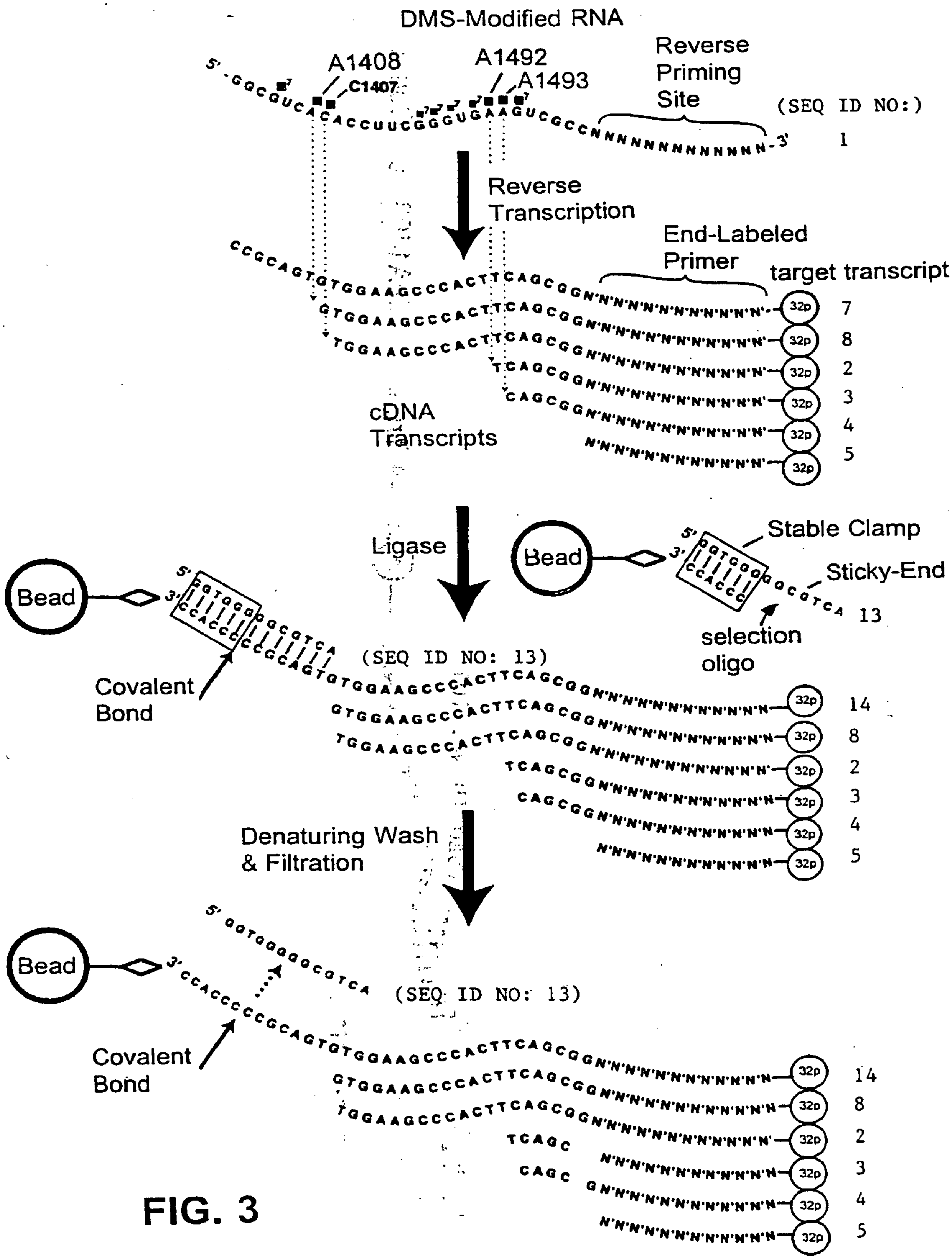

DMS-Modified RNA
A1408
C1407
A1492
A1493
Reverse Priming Site
(SEQ ID NO:)
1
Reverse Transcription
End-Labeled Primer
target transcript
7
8
2
3
4
5
cDNA Transcripts
Ligase
Bead
Stable Clamp
Sticky-End
13
selection oligo
Covalent Bond
(SEQ ID NO: 13)
14
Denaturing Wash & Filtration
(SEQ ID NO: 13)

FIG. 3

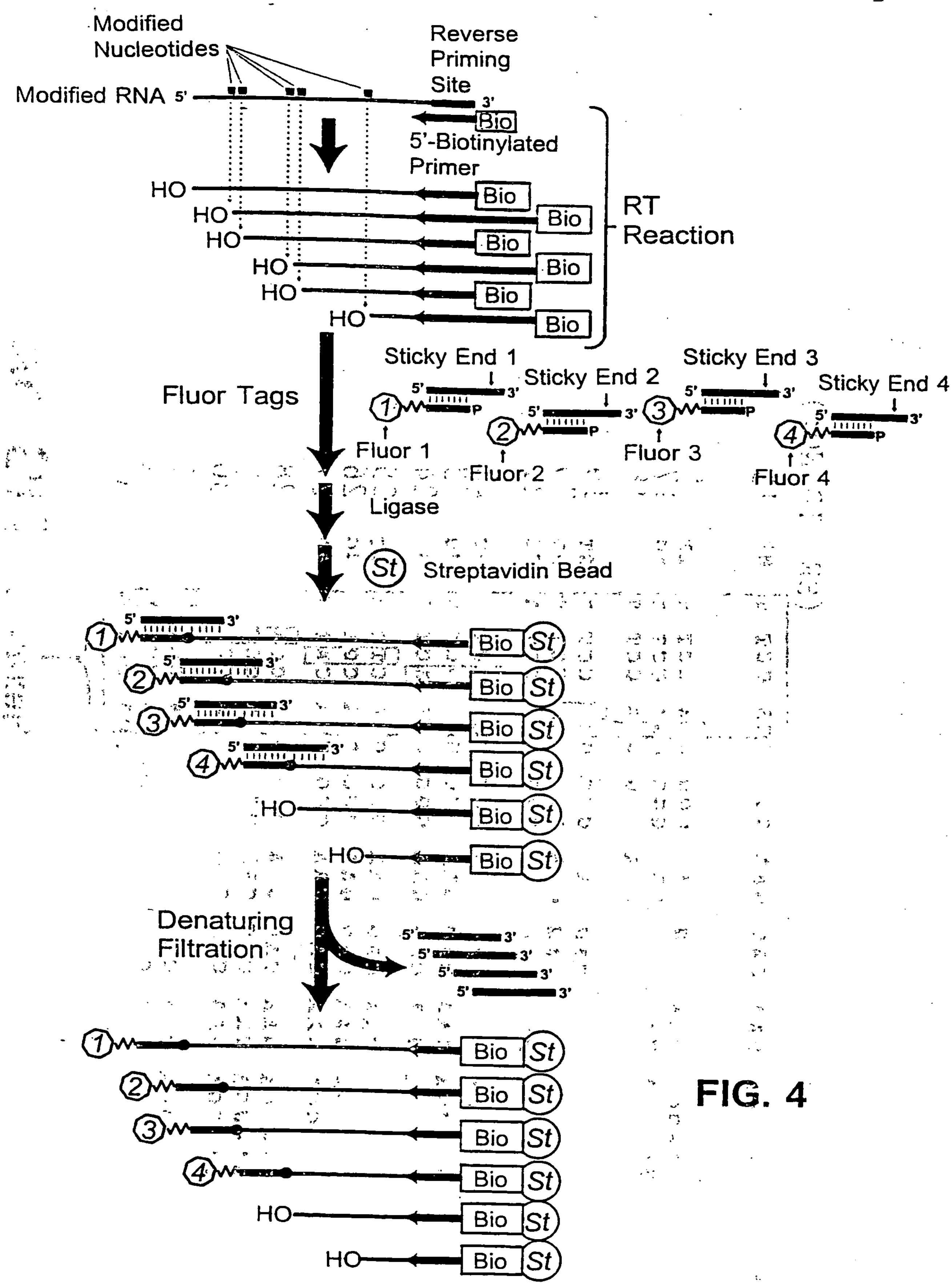

Modified Nucleotides
Reverse Priming Site
Modified RNA 5'
3'
Bio
5'-Biotinylated Primer
HO
Bio
RT Reaction
Fluor Tags
Sticky End 1
Sticky End 2
Sticky End 3
Sticky End 4
5'
3'
P
Fluor 1
Fluor 2
Fluor 3
Fluor 4
Ligase
St
Streptavidin Bead
Denaturing Filtration

FIG. 4

	Sequence		(SEQ ID NO:)
top	5'-GGA GAA CAG GAA GGG GCA CCT T	WT	15
	5'-GGA GAA CAG GAA GGG GTA CCT T	1T	16
	5'-GGA GAA CAG GAA GGG GGA CCT T	1G	17
	5'-GGA GAA CAG GAA GGG GAA CCT T	1A	18
	5'-GGA GAA CAG GAA GGG GCT CCT T	2T	19
	5'-GGA GAA CAG GAA GGG GCG CCT T	2G	20
	5'-GGA GAA CAG GAA GGG GCC CCT T	2C	21
	5'-GGA GAA CAG GAA GGG GCA TCT T	3T	22
	5'-GGA GAA CAG GAA GGG GCA GCT T	3G	23
	5'-GGA GAA CAG GAA GGG GCA ACT T	3A	24
	5'-GGA GAA CAG GAA GGG GCA CTT T	4T	25
	5'-GGA GAA CAG GAA GGG GCA CGT T	4G	26
	5'-GGA GAA CAG GAA GGG GCA CAT T	4A	27
	5'-GGA GAA CAG GAA GGG GCA CCA T	5A	28
	5'-GGA GAA CAG GAA GGG GCA CCC T	5C	29
	5'-GGA GAA CAG GAA GGG GCA CCG T	5G	30
	5'-GGA GAA CAG GAA GGG GCA CCT A	6A	31
	5'-GGA GAA CAG GAA GGG GCA CCT C	6C	32
	5'-GGA GAA CAG GAA GGG GCA CCT G	6G	33

Hexamer

FIG. 5A

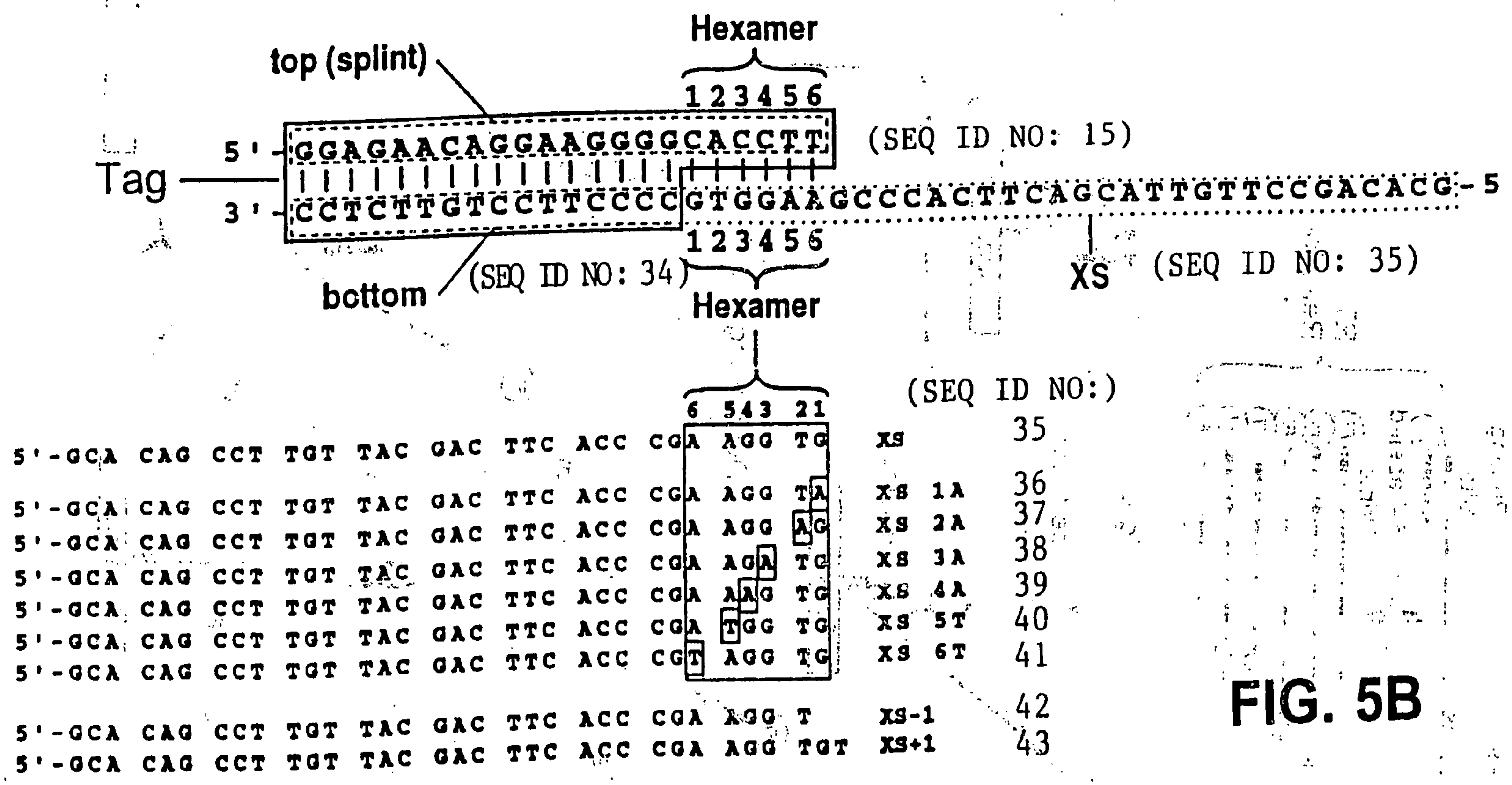

Hexamer
top (splint)
1 2 3 4 5 6
Tag
5'-GGAGAACAGGAAGGGGCACCTT (SEQ ID NO: 15)
3'-CCTCTTGTCCTTCCCCGTGGAAGCCCACTTCAGCATTGTTCCGACACG-5'
1 2 3 4 5 6
bottom
(SEQ ID NO: 34)
Hexamer
XS
(SEQ ID NO: 35)
(SEQ ID NO:)
6 543 21
5'-GCA CAG CCT TGT TAC GAC TTC ACC CGA AGG TG XS 35
5'-GCA CAG CCT TGT TAC GAC TTC ACC CGA AGG TA XS 1A 36
5'-GCA CAG CCT TGT TAC GAC TTC ACC CGA AGG AG XS 2A 37
5'-GCA CAG CCT TGT TAC GAC TTC ACC CGA AGA TG XS 3A 38
5'-GCA CAG CCT TGT TAC GAC TTC ACC CGA AAG TG XS 4A 39
5'-GCA CAG CCT TGT TAC GAC TTC ACC CGA TGG TG XS 5T 40
5'-GCA CAG CCT TGT TAC GAC TTC ACC CGT AGG TG XS 6T 41
5'-GCA CAG CCT TGT TAC GAC TTC ACC CGA AGG T XS-1 42
5'-GCA CAG CCT TGT TAC GAC TTC ACC CGA AGG TGT XS+1 43

FIG. 5B

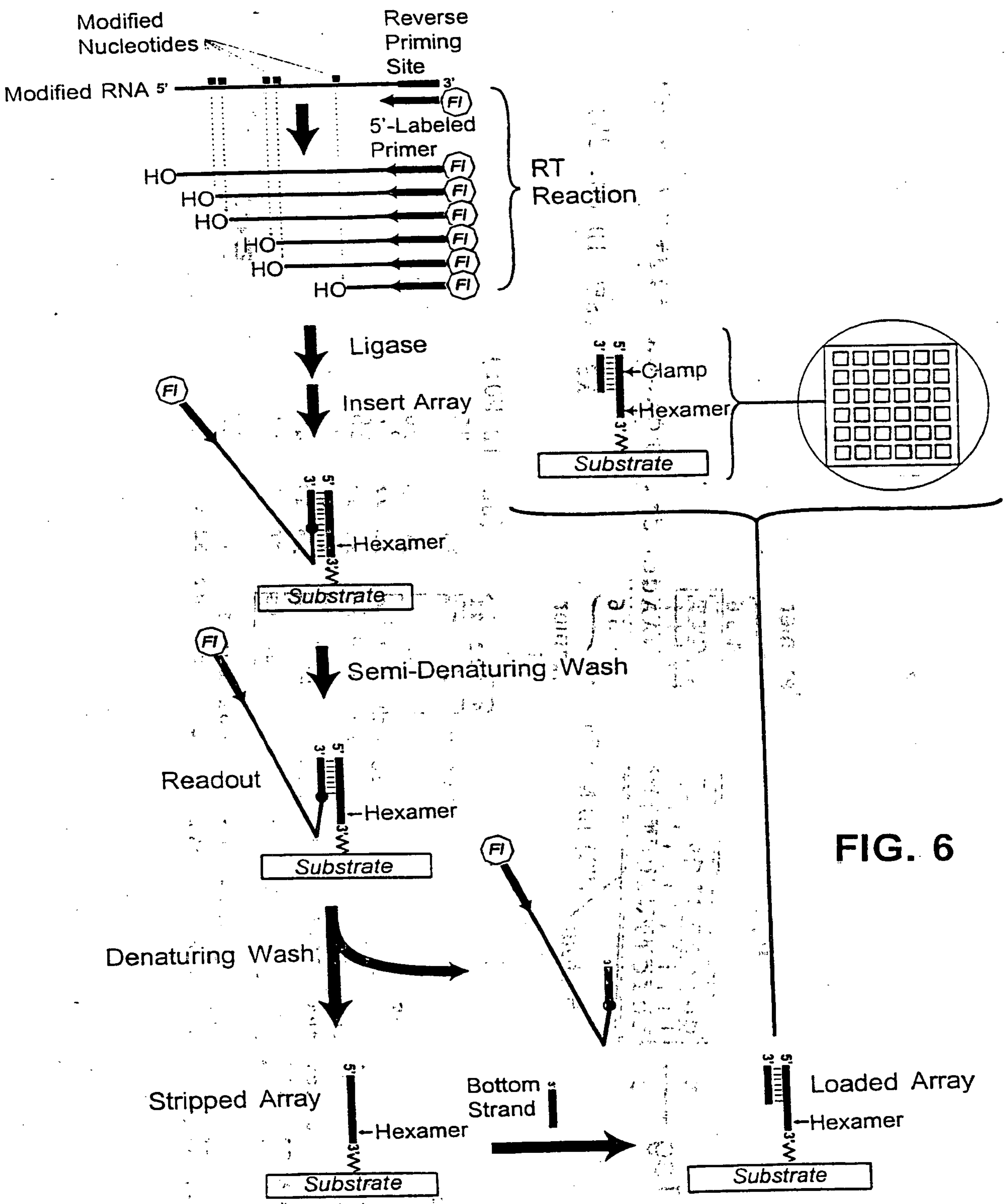

Modified
Nucleotides
Reverse
Priming
Site
Modified RNA 5'
3'
Fl
5'-Labeled
Primer
HO
RT
Reaction
Ligase
Insert Array
Clamp
Hexamer
Substrate
Semi-Denaturing Wash
Readout
Denaturing Wash
Stripped Array
Bottom
Strand
Loaded Array

FIG. 6

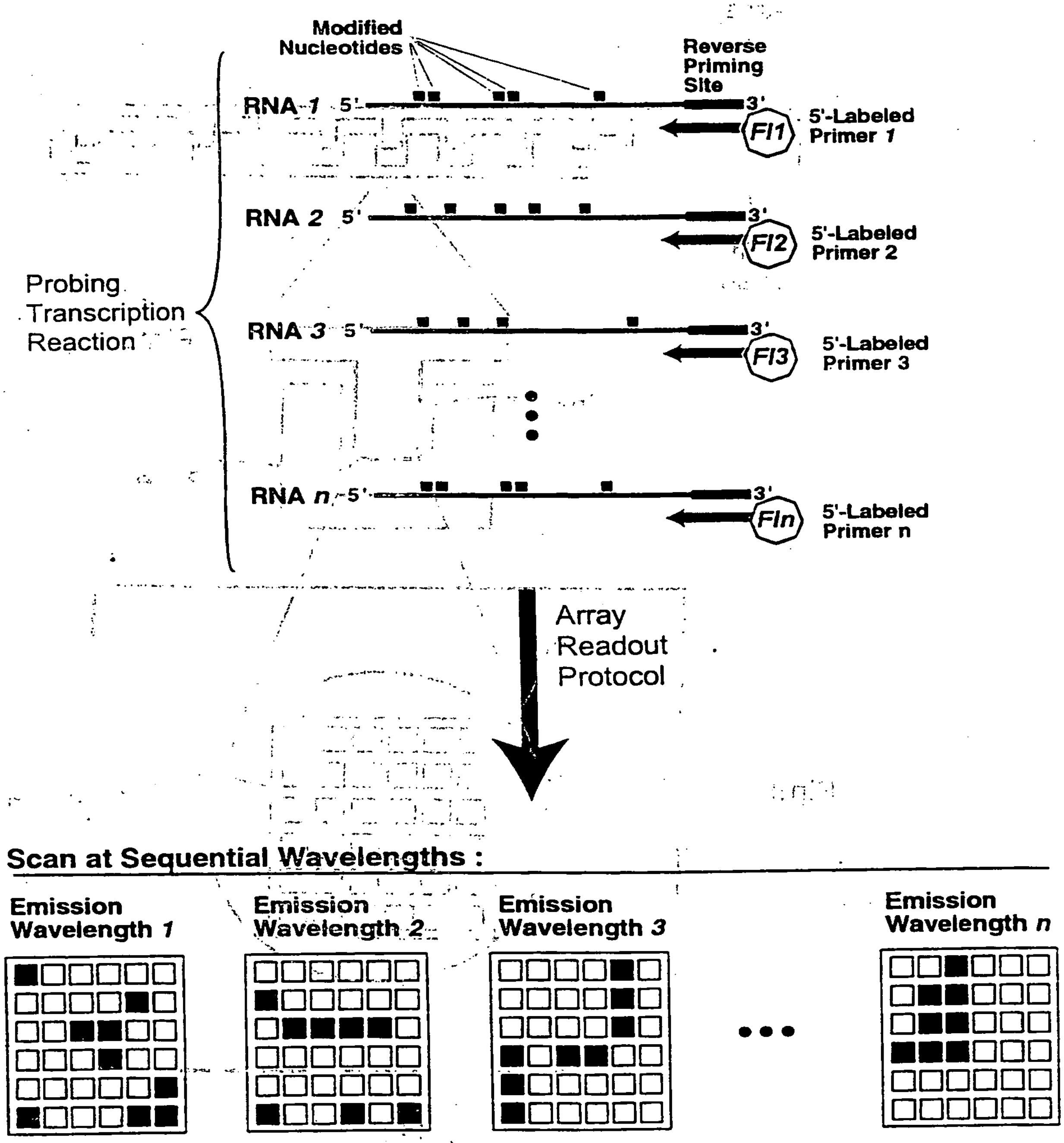

Modified Nucleotides
Reverse Priming Site
RNA 1 5' 3'
5'-Labeled Primer 1
Fl1
RNA 2 5' 3'
5'-Labeled Primer 2
Fl2
RNA 3 5' 3'
5'-Labeled Primer 3
Fl3
RNA n 5' 3'
5'-Labeled Primer n
Fln
Probing Transcription Reaction
Array Readout Protocol
Scan at Sequential Wavelengths :
Emission Wavelength 1
Emission Wavelength 2
Emission Wavelength 3
Emission Wavelength n

FIG. 7

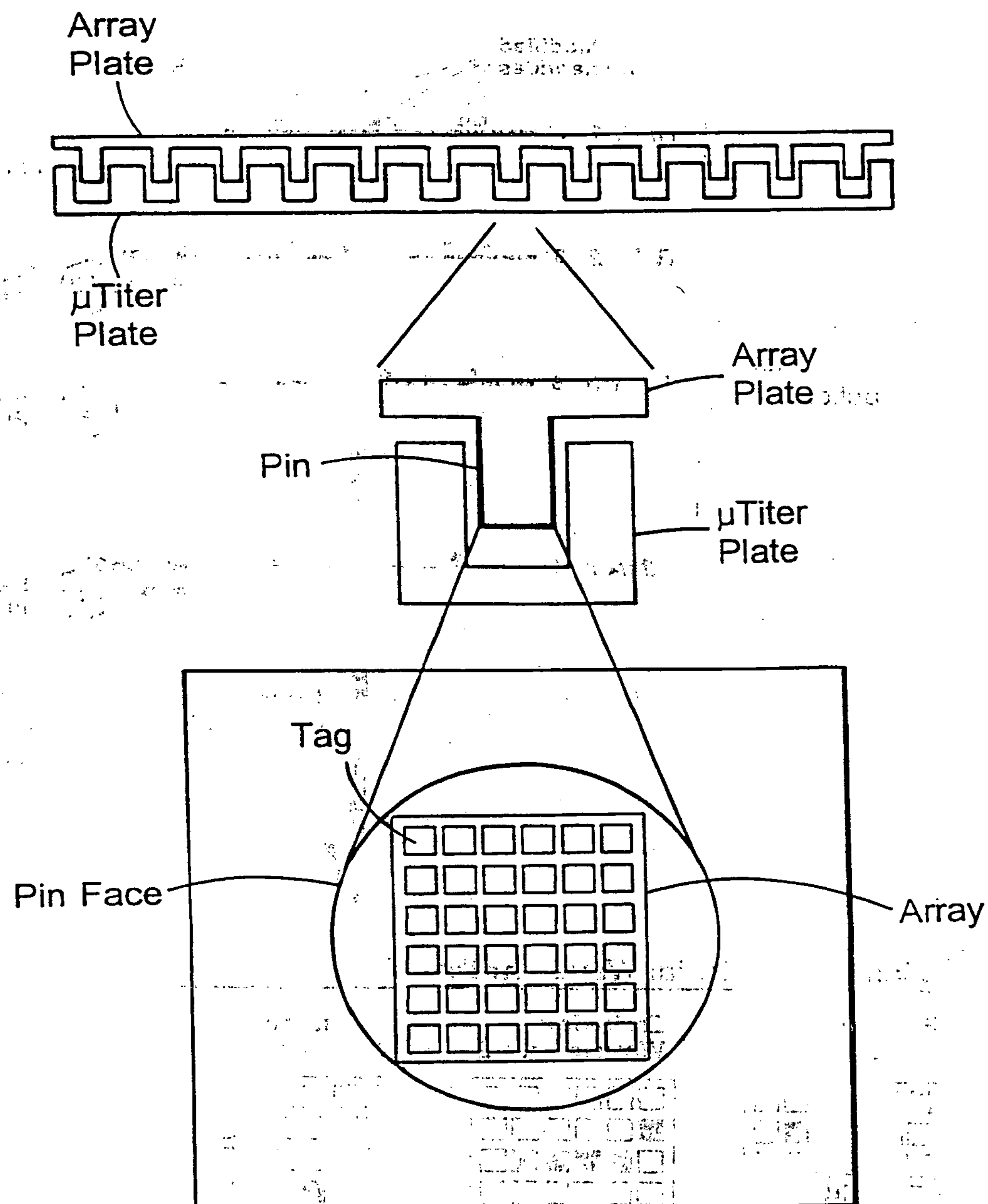
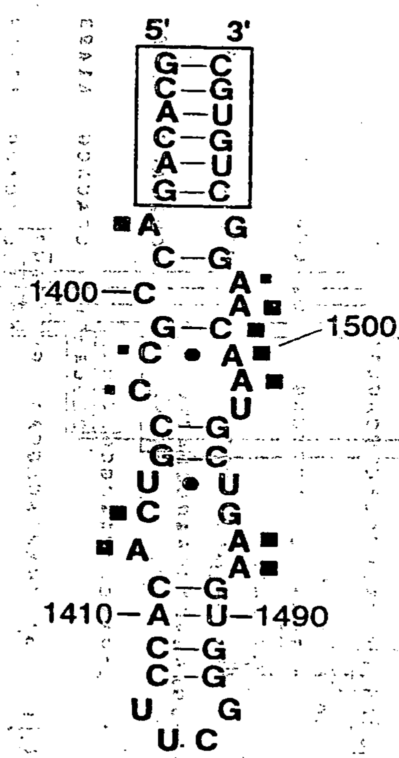

Array
Plate
µTiter
Plate
Array
Plate
Pin
µTiter
Plate
Tag
Pin Face
Array

FIG. 8

DMS METHYLATED NUCLEOTIDE: ■



(SEQ ID NO: 44)

FIG. 9A

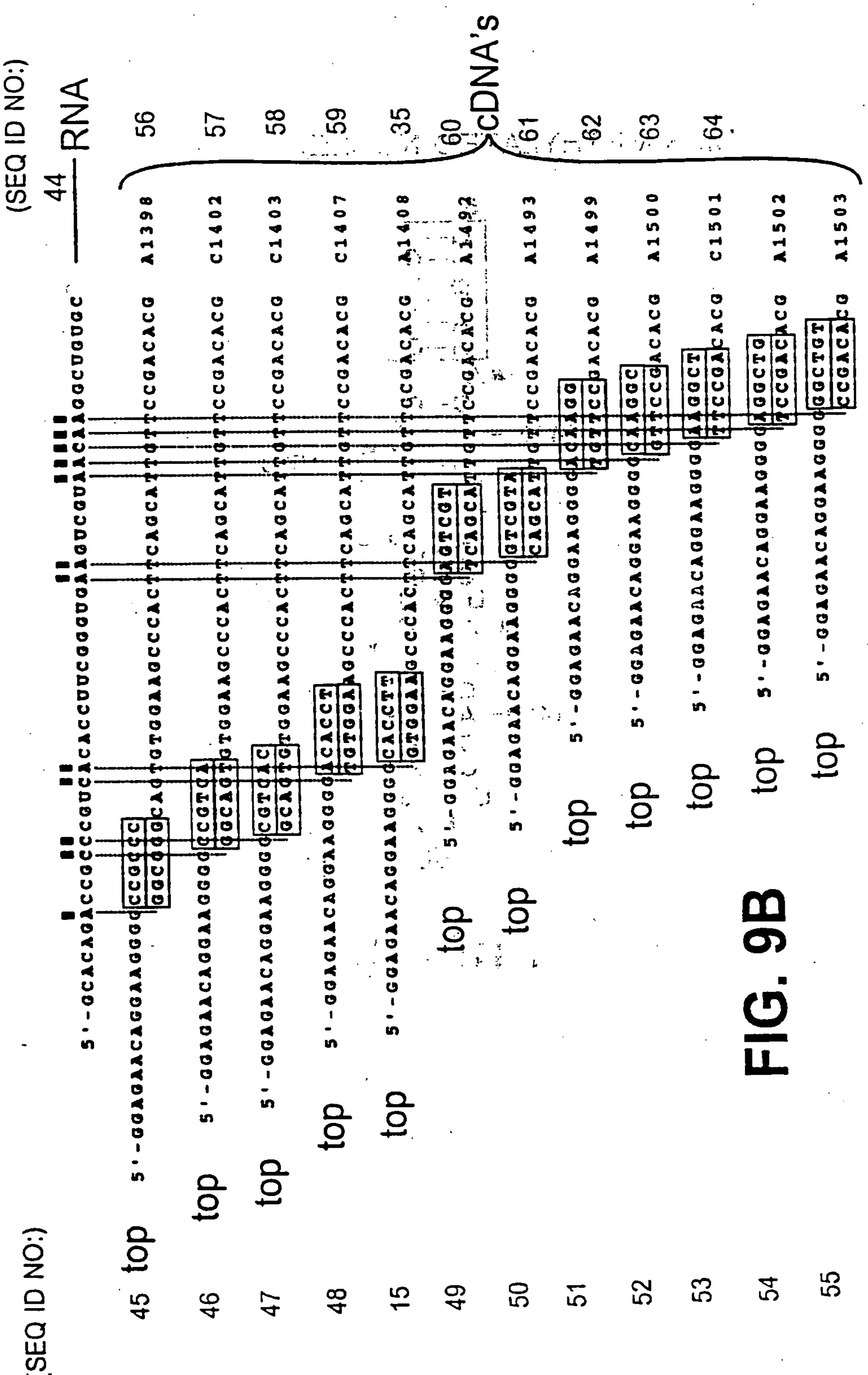

(SEQ ID NO:)
44
RNA
56
57
58
59
35
60
cDNA's
61
62
63
64
A1398
C1402
C1403
C1407
A1408
A1492
A1493
A1499
A1500
C1501
A1502
A1503
top
SEQ ID NO:)
45
46
47
48
15
49
50
51
52
53
54
55

FIG. 9B

3'-End Labeled Target

FIG. 10A

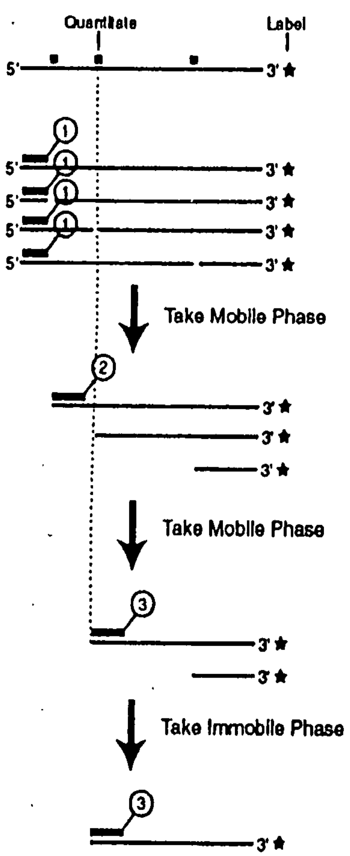


FIG. 10B

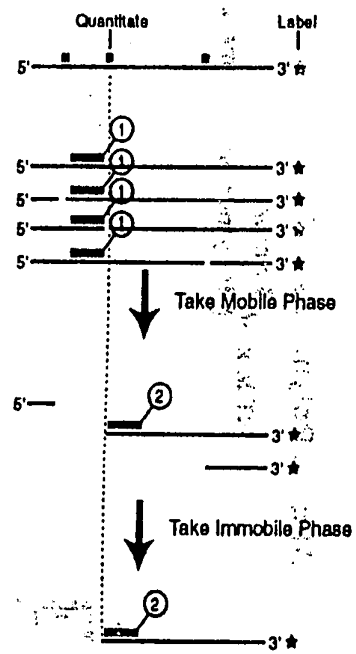


FIG. 10C

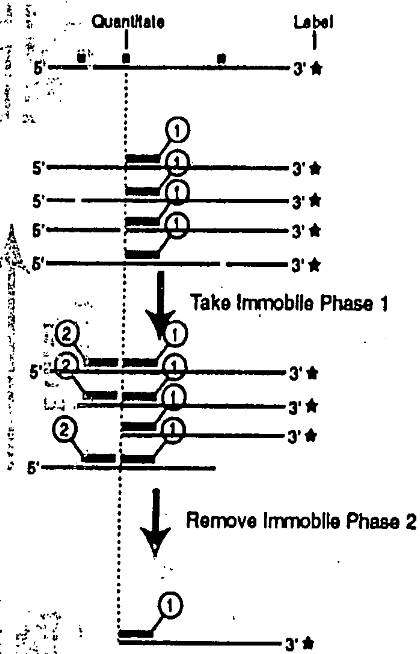
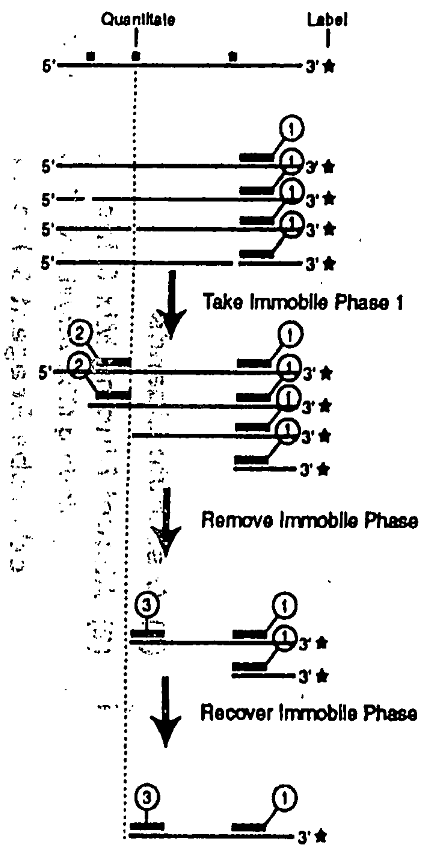


FIG. 10D



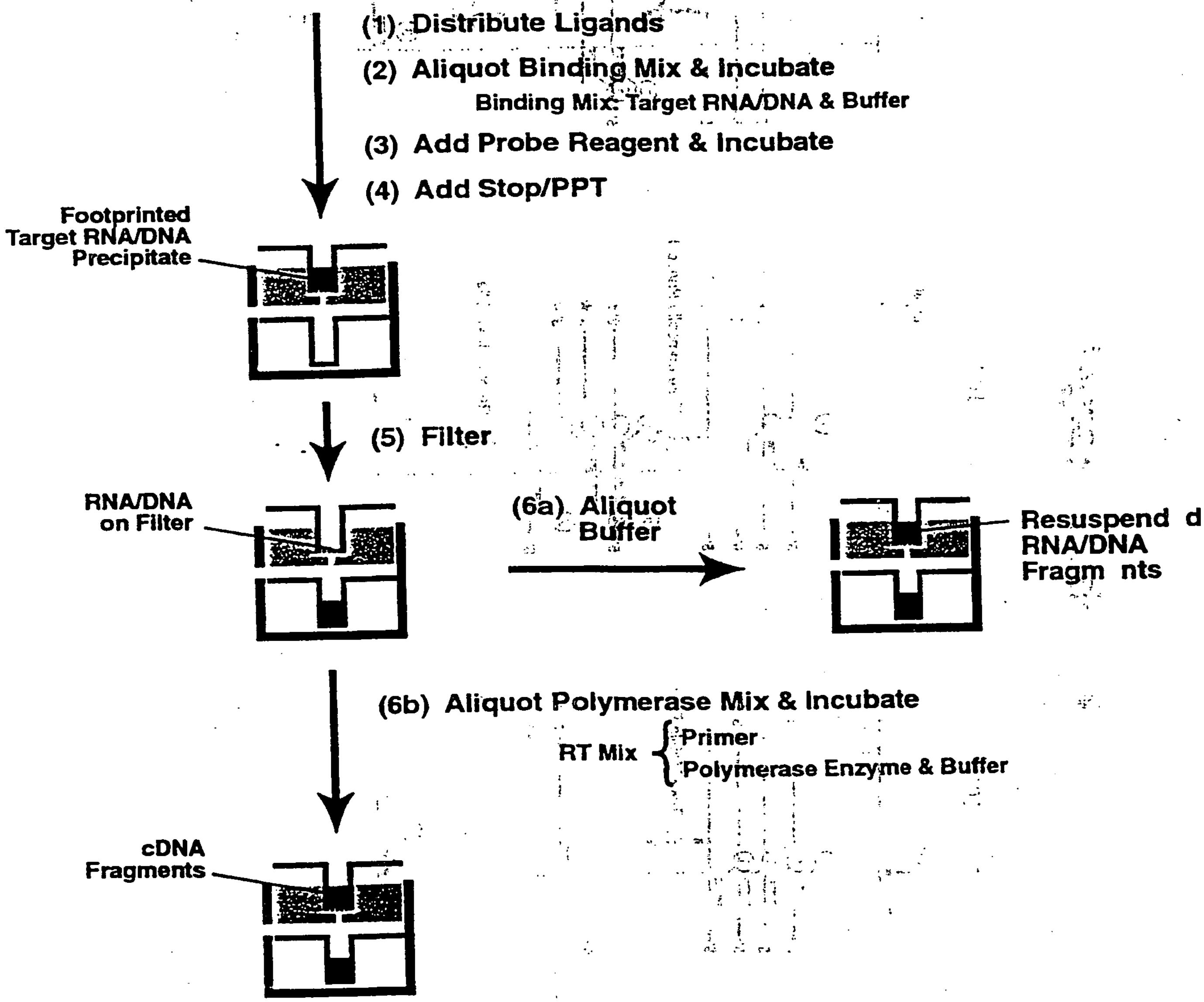

(1) Distribute Ligands
(2) Aliquot Binding Mix & Incubate
Binding Mix: Target RNA/DNA & Buffer
(3) Add Probe Reagent & Incubate
(4) Add Stop/PPT
Footprinted Target RNA/DNA Precipitate
(5) Filter
RNA/DNA on Filter
(6a) Aliquot Buffer
Resuspend d RNA/DNA Fragm nts
(6b) Aliquot Polymerase Mix & Incubate
RT Mix
Primer
Polymerase Enzyme & Buffer
cDNA Fragments

FIG. 11

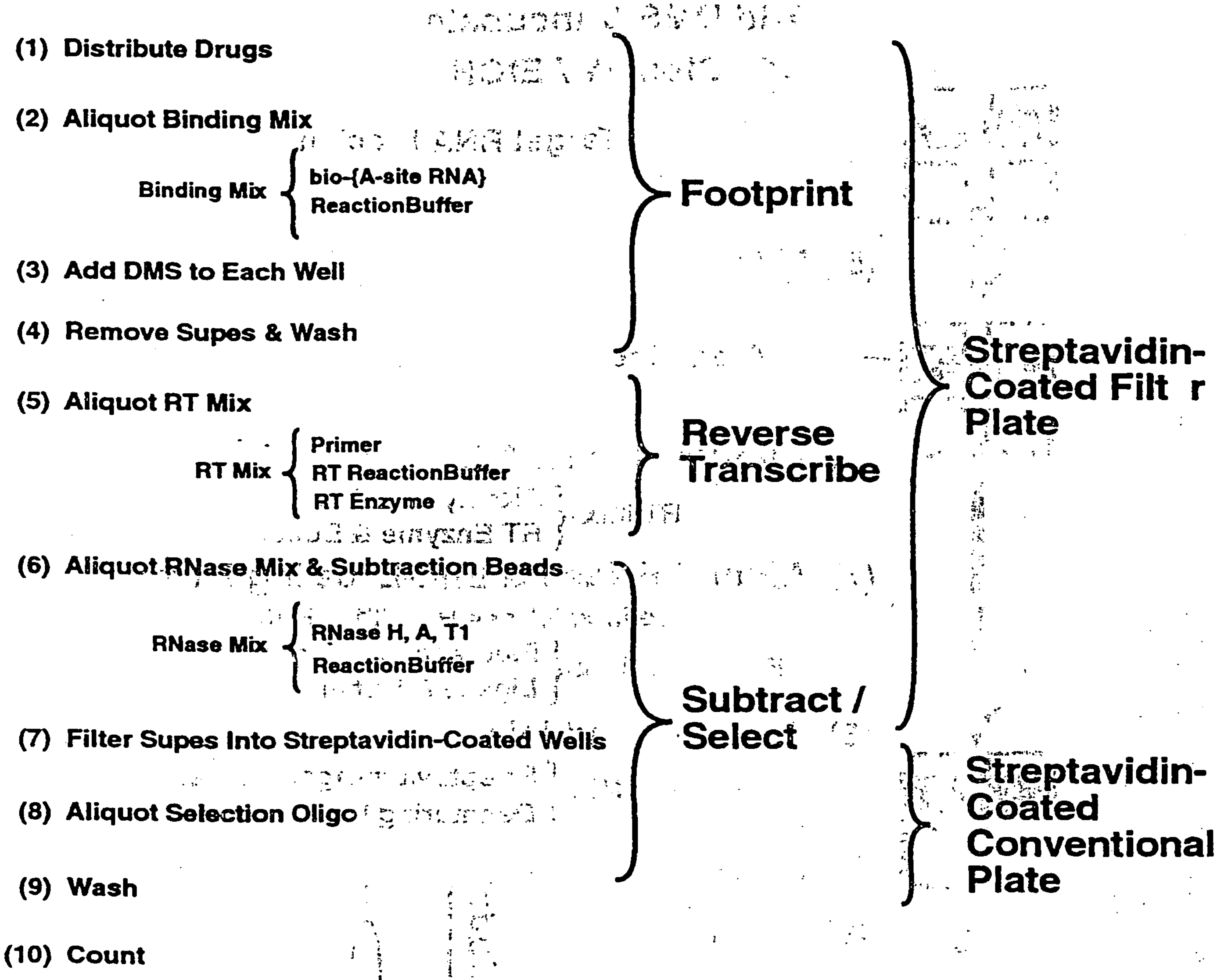

(1) Distribute Drugs
(2) Aliquot Binding Mix
Binding Mix
bio-{A-site RNA}
ReactionBuffer
(3) Add DMS to Each Well
(4) Remove Supes & Wash
Footprint
(5) Aliquot RT Mix
RT Mix
Primer
RT ReactionBuffer
RT Enzyme
Reverse Transcribe
(6) Aliquot RNase Mix & Subtraction Beads
RNase Mix
RNase H, A, T1
ReactionBuffer
(7) Filter Supes Into Streptavidin-Coated Wells
(8) Aliquot Selection Oligo
(9) Wash
(10) Count
Subtract / Select
Streptavidin-Coated Filt r Plate
Streptavidin-Coated Conventional Plate

FIG. 12

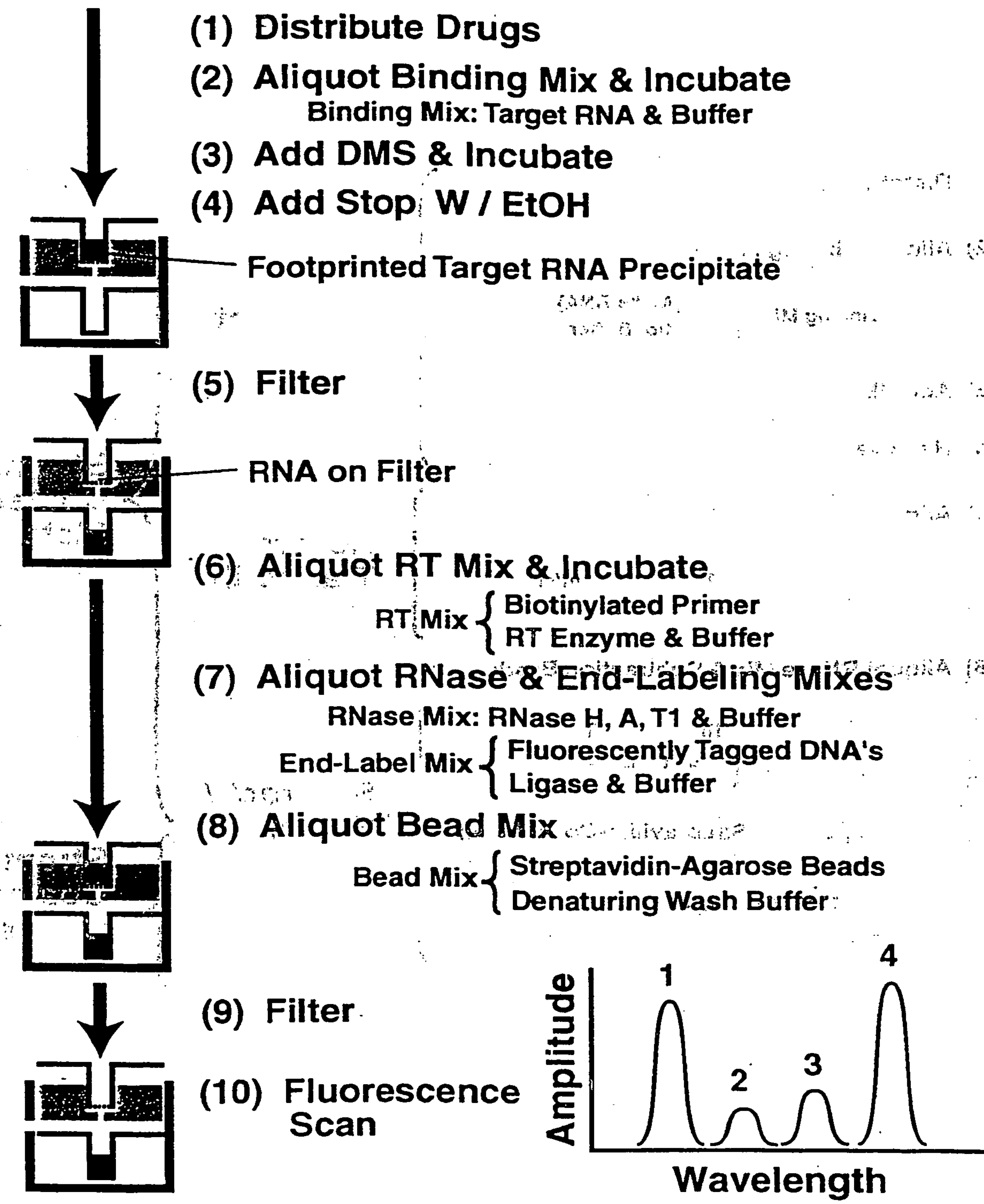

(1) Distribute Drugs
(2) Aliquot Binding Mix & Incubate
Binding Mix: Target RNA & Buffer
(3) Add DMS & Incubate
(4) Add Stop W / EtOH
Footprinted Target RNA Precipitate
(5) Filter
RNA on Filter
(6) Aliquot RT Mix & Incubate
RT Mix { Biotinylated Primer
RT Enzyme & Buffer
(7) Aliquot RNase & End-Labeling Mixes
RNase Mix: RNase H, A, T1 & Buffer
End-Label Mix { Fluorescently Tagged DNA's
Ligase & Buffer
(8) Aliquot Bead Mix
Bead Mix { Streptavidin-Agarose Beads
Denaturing Wash Buffer
(9) Filter
(10) Fluorescence Scan
Amplitude
Wavelength
1
2
3
4

FIG. 13

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/16211

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :C12Q 1/68; C07H 19/00, 21/00
US CL :435/6; 536/22.1, 25.32, 25.4
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6; 536/22.1, 25.32, 25.4

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS. STM
search terms: isolating, quantitiating, nucleic acid, hybridzation, modification, label

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	MATTHEWS et al. Analytical Strategies for the Use of DNA Probes. Analytical Biochemistry. 1988, Vol. 169, pages 1-25, see entire document.	1-10
Y	SYVANEN et al. Fast Quantitation of Nucleic Acid Hybrids by Affinity-Based Hybrid Collection. Nucleic Acids Research. 1986, Vol. 14, No. 12, pages 5037-5048, see entire document.	1-10

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
01 OCTOBER 1998	28 OCT 1998

Name and mailing address of the ISA/US	Authorized officer
Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231	JEZIA RILEY
Facsimile No. (703) 305-3230	Telephone No. (703) 308-0196

INTERNATIONAL SEARCH REPORT	International application No. PCT/US98/16211

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING
This ISA found multiple inventions as follows:

Group I, claim(s)1-10, drawn to method of isolating a target nucleic acid.

Group II, claim(s) 11-25, drawn to method of quantitating a target nucleic acid.

The inventions listed as Groups I and II do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: the invention of group I do not require the technical feature of the invention of Group II.